UiO : Department of Mathematics
University of Oslo

The Poisson-Binomial Model for Fish Abundance Estimation

With Applications to Northeast Arctic Cod

Kjersti Moss

Master's Thesis for the degree
Modelling and Data Analysis (MOD5960), November 2015

Abstract

The Institute of Marine Research collects data from different sources for the estimation of fish abundance. These data can be divided into two groups:

1. Data from research surveys.

2. Fishery based data.

In this thesis, we aim to utilize both data sets to estimate the abundance of fish, along with the catch. In addition to a point estimate, we wish to assess the uncertainty in these estimates. More formally, we hope to find quantiles of the simultaneous distribution

$$\pi(\mathbf{N}, \mathbf{C} | \mathbf{D}^N, \mathbf{D}^C, \text{parameters}),$$

that is, the distribution of the abundances and catches, given both data sources.

On the way towards this goal, we need to specify a model for the abundance, the catch, and the data. The model for the abundance and catch is what we call *the Poisson-binomial model*. This model is the central theme of the thesis. We explore the properties of the model, and derive conditions for when it is identifiable. Furthermore, we investigate both a frequentistic and a Bayesian method to estimate the model parameters. It turns out that we are not able to describe the simultaneous distribution $\pi(\mathbf{N}, \mathbf{C} | \mathbf{D}^N, \mathbf{D}^C, \text{parameters})$ analytically, and we can neither sample directly from it. However, we can obtain Monte Carlo samples of this distribution through importance sampling techniques, and thereby calculate approximate quantiles.

The methods we develop are applied to data on Northeast Arctic cod (Skrei in Norwegian), from the years 1985-2003.

Acknowledgements

First of all, I thank my main supervisor Geir Storvik for his guidance and enthusiasm for this project. I also thank my co-supervisor Sondre Aanes for sharing his insight into the fishery science. Further, I must thank the best mathematics teachers from my school years, Ingve Tofteland and Tor Arne Mjølund, who inspired me to study the subject further. The data set I analyse was provided by the Norwegian Institute for Marine Research (IMR), and for that I thank them. I am also grateful to IMR and the Institute of Mathematics at the University of Oslo, that I was allowed to participate at a one-day seminar at IMR in Bergen.

When it comes to the writing process, I thank my brothers Rolf and Scott who have proofread much of my material, and my sister Elen who have been babysitting in the last hectic days.

Further, I thank the Norwegian welfare system, and especially *Lånekassen*, *Husbanken*, and *Sogn Studentbarnehage*. Without you it would have been both economically and practically impossible to complete this project.

The past year have been very pleasant socially, as well as academically. For that I thank my fellow master students, and everyone else on the 8'th floor of Niels Henrik Abels House, Blindern, Oslo.

Finally, I thank my husband Jonas Moss for a great collaboration through our master studies, and otherwise.

Kjersti Moss,
Oslo, November 16, 2015

Dedicated to my children
Richard and Ervin

Contents

Notation and definitions

For future reference.

$N_{y,a}$	The number of a years old fish alive at the beginning of year y.
$\overline{N}_{y,a}$	The number of a years old fish alive at the end of year y.
$C_{y,a}$	The number of a years old fish caught in year y.
$m_{y,a}$	The natural mortality, which is the probability of dying of natural causes for an a years old fish in year y.
$f_{y,a}$	The fishery mortality, which is the probability of being fished for an a years old fish in year y.
$I_{y,a}$	An index containing information about the number of fish.
q_a	The catchability of the survey.
$\mathbf{D}^N, \mathbf{D}^C$	The available data. $\mathbf{D}^N = \{I_{y,a}\}$, and $\mathbf{D}^C$ is the fishery based data.
Cohort	A cohort is a group of animals of the same species, identified by a common characteristic. In this thesis, the common characteristic is the year of birth.
Population	A population is a summation of all the organisms of the same group or species, which live in a particular geographical area, and have the capability of interbreeding.
Recruitment	Recruitment is the number of new young fish that enter a population in a given year. Those individuals are often called the *recruits*.

Figure 1: The Norwegian *Skrei* is popular in many countries. This is an advertisement from the German food recipe portal fischausnorwegen.de [2]. *Alle guten Dinge sind Skrei*, plays on the expression *Aller guten Dinge sind drei*. This is also a common expression in English, namely *All good things come in threes*.

[2]http://fischausnorwegen.de/Kampagne/Alle-guten-Dinge-sind-Skrei-2013

Chapter 1

Introduction

> Skuld' torsken os feile, hvad havde vi da!
>
> Petter Dass in *Nordlands Trompet*

Every year the coast of Northern Norway is visited by millions of spawning *skrei*, which is a population of cod. The coast of the Lofoten Islands is the honeymoon destination for most of them, and inhabitants of Lofoten have enjoyed the food the skrei provide for many centuries. The large abundance of skrei allowed fishermen to be well fed, and still have a surplus to trade with other goods. Already in the 11th century trading routes to the south of Norway were established, and in the 12th century skrei was shipped to England as well. From the mid 14th century the Hanseaten controlled most of these trading routes, and the Germans also became important consumers.

We need to mention an important invention that made the distribution of skrei possible: The stockfish [1]. Stockfish is what we call the skrei after it has been dried on a *drying flake* for about three months, and then indoor for another two-three months. After this process the durability of the fish is extended by many years. Thus it was possible to export it over long distances, at any time of the year.

Today there are more modern methods of food processing, but still stockfish is made in Lofoten and sold to many corners of the world. Two pictures from drying flakes outside Svolvær are displayed in Figure 1.1.

After a millennium the skrei is still important for the Norwegian economy. In 2013 the export value exceeded one billion NOK for the first time, which is about 122 million U.S.

[1] The Norwegian word for cod, *torsk*, is derived from the Old Norse word *thorskr*, akin to the word *tørr* (dry in English). So "torsk" directly refers to stockfish (tørrfisk)! http://www.wordreference.com/definition/torsk

Figure 1.1: Northeast Arctic cod. Pictures from drying flake in Svolvær, Lofoten, Norway. Summer 2014. Photo credit: Kjersti Moss

dollars (Jakobsen and Ozhigin, 2011). However, money is not everything. Other positive features of this industry includes that the resource is renewable, the food it delivers is healthy, and it employs many people. This is something worth protecting, and what the industry needs protection form is itself. Protection from overfishing. If the fish stock collapse, it might never rise again.

This is where the estimation of fish abundance, which is the topic of this thesis, becomes relevant. Biologists have quite good ideas about what *percentage* of the cohorts that can be fished, without risking a collapse. Let us say this level is at 25 percent. Now the Norwegian government can not make a law that states "this year, fishermen are allowed to fish every 4th skrei they see - and they shall make sure not to see the same fish twice!". That would of course be absurd. But if we can find a good estimate of how large the stock is, then we can make a law that states how many tonnes fishermen are allowed to fish. This is measurable and possible to enforce. It is well known that such laws already exists, and we know them as fishing quotas, or Total Allowed Catch (TAC). The TAC is set every year, and is based on *The Harvest control rule*, which was designed by The Joint Norwegian-Russian Fishery Commission in the autumn of 2002 (ICES, 2014). The Harvest control rule states, among other things, that:

> "[...] the management strategies for cod should take into account [...] :
> *full utilization of all available information on stock development.*"

A full utilization of the available data requires good statistical models, which motivates us to search for models than are better than those used today. In this thesis we investigate if

the *Poisson-binomial model* might be such a model.

1.1 Northeast Arctic cod

Chapter 5.4 in the book *The Barents Sea: Ecosystem, Resources, Management* by Jakobsen and Ozhigin (2011) is used as reference for this entire section.

The methods that we will develop in this thesis aim to be general and applicable to all populations of fish, but we will focus specifically on skrei. Thus we should know a little bit about this population of cod.

Skrei is the Norwegian name for the migrating cod in the Barents and North Seas. This is the largest population of the species *Gadus Morhua*, also known as *the Atlantic cod*. The term cod actually refers to the taxonomic family *Gadidae*, which contains 13 different genera, one of which is *Gadus*. Thus we try to avoid just writing "cod" when we refer to skrei. Henceforth we will use the term *Northeast Arctic cod*, abbreviated by NEA cod, and leave our Norwegian word behind. NEA cod is the term used in both Jakobsen and Ozhigin (2011) and ICES (2014). Previously the term *Arcto-Norwegian cod* was commonly used, and still many other names are used. Among these are *Barents Sea cod*, *Lofoten cod* and *Lofoten-Barents Sea cod*.

The NEA cod is not only the largest population of Atlantic cod, but the largest stock of any cod. The ICES estimate of the population size was about 1.2 billion in 2014. If we multiply this by the average weight of 6 kg, we get about 7.2 billion kg - or 7.2 million tonnes. That is more than one tonne of NEA cod per Norwegian citizen. Or almost 3 kilos per Norwegian per day of a year. As previously mentioned we can not harvest the whole stock, but this illustrates that the NEA cod can feed a big part of Norway's population alone.

In addition to being an important commercial fish, the NEA cod dominates the Barents Sea ecosystem. It is a carnivore that feed on everything from zooplankton to small fishes, including young NEA cod, during its lifetime. The feeding grounds in the Barents Sea areas are marked in Figure 1.2, together with the spawning grounds.

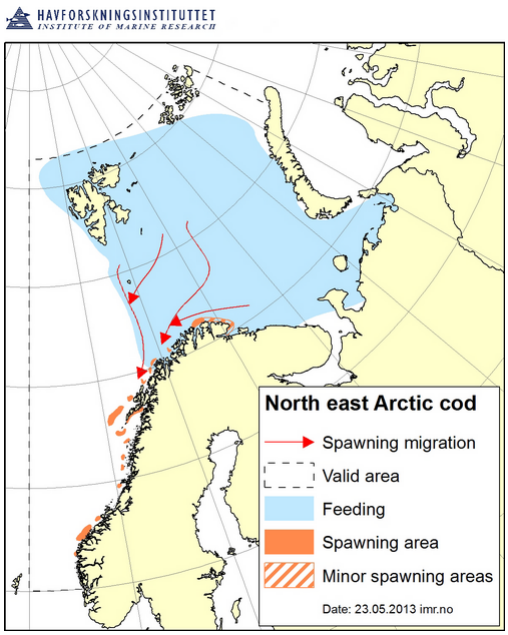


Figure 1.2: The blue area is the extent of occurrence, while the orange areas are the spawning grounds. The right arrows indicate the spawning migration. Figure credit: The Norwegian Institute of Marine Research.

The reproduction

In January the spawning stock migrates to the coast of Norway to reproduce. This happens every year after the fish is aged from four to six years. The spawning period lasts until March, when the NEA cod returns to its Barents Sea feeding grounds. A spawning cod can migrate more than 2000 km during one year.

The NEA cod is a batch spawner and there has been registered three to eight batches per season. In each batch, 60-310 thousand eggs are released from each female. So like most other fishes, the NEA cod bets on quantity rather than quality. Only 2-3 percent of the eggs reach the stage of first-feeding larvae, and only about 0.00065 percent of the eggs live until recruitment age at 3 years.

Many factors influence how many eggs that survive. One factor that has been investigated is the temperature, where a positive correlation was found. This is probably caused by increased levels of zooplankton for the larvae to feed on.

1.2 Current stock assessment

Every year the International Council of the Exploration of the Sea (ICES) publish a large report on the status of many fish populations in the North Sea, and the NEA cod is one of them. The report form 2014 ICES (2014) is a reference for the following section.

When the Arctic Fishery Working Group (AFWG), one of the ICES expert groups, sits down to determine the status of the NEA cod, they have two main sources of data.

Fishery based data, $\mathbf{D}^C$

The fisheries are obliged to report the weight of their catch. In addition to this, from some boats the length of individual fish is measured and for some even age is measured. Estimates of the catches $\{C_{y,a}\}_{y=1,a=1}^{Y,A}$, for every year and every age group, are made from these data.

Data from research surveys, $\mathbf{D}^N$

Every year a research vessel trawls given areas of the Barents Sea. The trawls are weighted and, like for the catch data, length and age are measured for some individuals. These data are used to obtain age-year specific indices $\mathbf{D}^N = \{I_{y,a}\}_{y=1,a=1}^{Y,A}$. These indices are assumed to be estimates of $\{q_a N_{y,a}\}_{y=1,a=1}^{Y,A}$, where $\{N_{y,a}\}_{y=1,a=1}^{Y,A}$ are the abundances in year y for age group a, and the parameter q_a is a catchability that may vary with age. Some age groups may be easier to get into the trawl than others.

The stock assessment is made by a kind of Virtual Population Analysis (VPA) model, called Extended Survivor Analysis (XSA). The VPA model is described e.g. in Lassen and Medly (2001). To use the VPA, one needs to assume the catches $\{C_{y,a}\}_{y=1,a=1}^{Y,A}$ to be known for every year-age group.

Further, one needs some starting values of the abundances $\{N_{Y,a}, N_{y,A}\}_{y=1,a=1}^{A,Y-1}$ in the last year and last age group. The abundance in the last year Y is of course unknown, and is the quantity we are most interested in. However, different values can be tried out, and then be compared with the indices $\{I_{y,a}\}$.

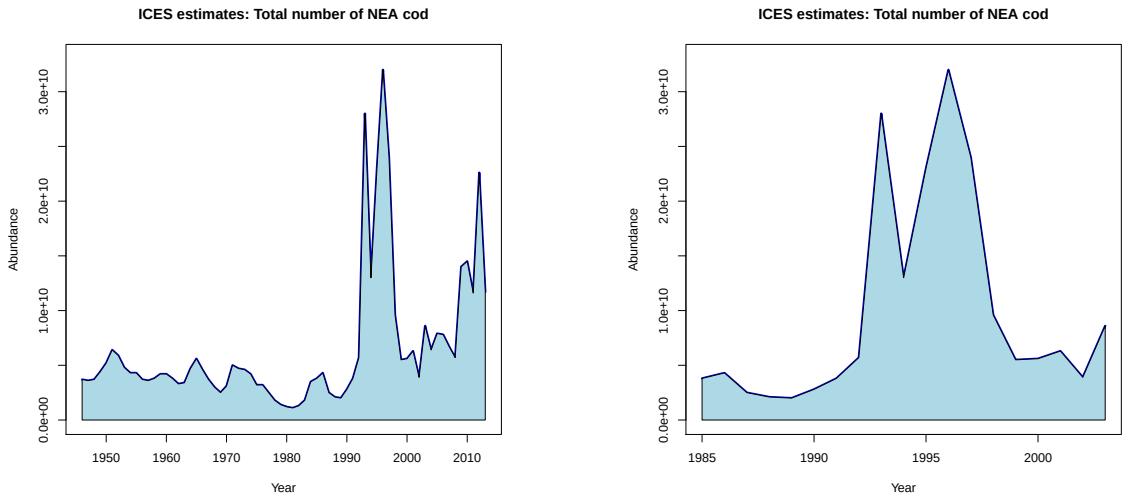
Finally, the VPA requires some values for the natural *instantaneous mortalities* $\{M_{y,a}\}_{y=1,a=1}^{Y,A}$, which are distinguished from the fishery instantaneous mortalities $\{F_{y,a}\}_{y=1,a=1}^{Y,A}$. The fishery mortalities are estimated in the VPA. The instantaneous mortalities has the interpretation

that $1 - e^{-M_{y,a}}$ is the probability of dying of causes other than fishing in year y for fish in age group a.

Given the catches, starting values and instantaneous natural mortalities, the abundances and instantaneous fishery mortalities are deterministically calculated.

The ICES report have used a more complicated model that builds on this idea. This is called the Extended Survivor Analysis (XSA), and is described by Shepherd (1999). We will not go further into detail about this model here, but we should mention that the analysis assumes the natural mortality to be constant over age groups and over years, at $M = 0.2$. They also add a cannibalism mortality for the youngest age groups. The 0.2-assumption is debated, and other values have been suggested. In the mid '80s, Tretyak investigated a model of age-dependent mortalities, and arrived at an average mortality of $M = 0.12$. However, the 0.2-estimate remains the standard.

Now let us take a look at the results of the official analysis by the AFWG. Figure 1.3(a) shows the ICES estimates from 1946 to 2013. The plot is made in `R` (R Core Team, 2015), using the numbers from Table 3.16 in ICES (2014). We are also interested in taking a closer look at the years between 1985 and 2003, and this is shown in Figure 1.3(b). This is interesting because we will use data from this period in this thesis. The first thing we notice is a huge spike in the mid '90s. This can partly be explained by reduced fishing activity in the early '90s.



(a) From year 1946 to 2013.

(b) Zooming in on the years 1985-2003.

Figure 1.3: ICES estimates of the number of NEA cod. The graph is made by the author of this thesis, using the numbers from the ICES 2014 report (ICES, 2014).

The restrictions on the fishing fleets were enforced due to a fear of overfishing. If we look closely at the years before 1990, we might infer that there is a down-going trend. This looked particularly frightening at the time, because we had the knowledge of what was happening outside the coast of Canada. In 1992 the Northwest Atlantic cod stock size had dropped to only one percent of its earlier level (Hamilton and Butler, 2001). This collapse may be permanent, as a potential recovery has still not taken place in 2015. Thus one can argue that Norway's concerns were well justified. We do not know how the stock would have developed without the restrictions, but we know that the population thrived during this period. After a drop in the beginning of the '00s, the population is still very viable today.

1.3 Introduction to the thesis project

The aims of this project are described in the *Project description* in the Master's agreement. This section is a summary of the Project description, and an outline of the contents of the thesis. The ultimate goal is to obtain samples from the posterior distribution

$$\pi(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C), \tag{1.1}$$

where $\mathbf{D}^N$ and $\mathbf{D}^C$ are the survey based data and the fishery based data, described in the previous section. $\mathbf{N} = \{N_{y,a}\}$ is the set of all abundances, over every year and age group, and similarly, $\mathbf{C}$ is the set of all the catches.

As suggested in the Project description, we will not be able to sample directly from $\pi(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C)$. However, we will be able to sample from an approximate distribution $q(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C)$. Then we can obtain distributional properties of the abundance or catch, which we generally can describe as a function $h(\mathbf{N}, \mathbf{C})$, through importance sampling ideas:

$$\begin{aligned}
E^\pi[h(\mathbf{N}, \mathbf{C})|\mathbf{D}^N, \mathbf{D}^C) &= \sum_{\mathbf{N},\mathbf{C}} h(\mathbf{N}, \mathbf{C})\pi(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C) \\
&= \sum_{\mathbf{N},\mathbf{C}} h(\mathbf{N}, \mathbf{C})\frac{\pi(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C)}{q(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C)} q(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C) \\
&= E^q[h(\mathbf{N}, \mathbf{C})\frac{\pi(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C)}{q(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C)}] \\
&\approx \frac{1}{B}\sum_{b=1}^{B} h(\mathbf{N}^b, \mathbf{C}^b)\frac{\pi(\mathbf{N}^b, \mathbf{C}^b|\mathbf{D}^N, \mathbf{D}^C)}{q(\mathbf{N}^b, \mathbf{C}^b|\mathbf{D}^N, \mathbf{D}^C)}.
\end{aligned}$$

where $\{(\mathbf{N}^b, \mathbf{C}^b)\}$ are samples from $q(\mathbf{N}^b, \mathbf{C}^b|\mathbf{D}^N, \mathbf{D}^C)$. The last term in the sum above is

called the importance weight. In order to obtain these samples we face two main challenges (Project description):

1. To specify a realistic model $\pi(\mathbf{N}^b, \mathbf{C}^b | \mathbf{D}^N, \mathbf{D}^C)$.

2. To specify an approximate distribution $q(\mathbf{N}^b, \mathbf{C}^b | \mathbf{D}^N, \mathbf{D}^C)$ which both is easy to sample from and at the same time gives computable importance weights.

To answer the first challenge, we work further on a Poisson-binomial model that Geir Storvik has been working on earlier. In addition to presenting his ideas and results, we will justify and discuss the model assumptions and explore the properties of the model further. The basic idea is that the abundances in age group 1 are Poisson distributed, and the catches given the abundances are binomially distributed. That is

$$\begin{aligned} N_{y,a=1} &\sim \text{Poisson}(\lambda_{y,a=1}), \\ C_{y,a} | N_{y,a} &\sim \text{Binomial}(N_{y,a}, (1 - m_{y,a}) f_{y,a}), \end{aligned}$$

for all years $y \in (-A, ..., 1, ..., Y)$ and all ages $a \in (1, ..., A)$. Here, $m_{y,a}$ is the probability of dying of natural, or non-fishery, causes in year y for fish of age a. Further, $f_{y,a}$ is the probability of being fished, given that the fish has not died of other causes. This is explained carefully in Chapter 3, which is dedicated to the Poisson-binomial model.

We are not only interested in samples from the posterior distribution in (1.1). It is also interesting in itself to make some inference about the parameters in our model. That is; the catchabilities $\{q_a\}$, the probabilities of dying of non-fishery causes $\{m_{y,a}\}$ and the probabilities of being fished $\{f_{y,a}\}$. These probabilities relate to the instantaneous mortalities, which are used in the ICES (2014)-report, in the following way:

$$\begin{aligned} m_{y,a} &= 1 - e^{-M_{y,a}}, \\ f_{y,a} &= 1 - e^{-F_{y,a}}. \end{aligned}$$

To find out what is actually possible to estimate, we investigate whether our model is identifiable and/or estimable. That is, is it possible to make inference about all the parameters? If not all, then which? This is the theme of Chapter 4.

In Chapter 5 we present two different ways to make inference about the model parameters using the available data. We also describe importance sampling carefully, and how we use it to obtain samples of (1.1). An extension of importance sampling, called *sequential importance sampling with resampling* (SISR) is also presented and applied. The products of Chapter 5 are two `R` functions that output posterior samples of (1.1), given data (and a bit more!).

In Chapter 6 we perform a simulation experiment to investigate how our methods work. Further, in Chapter 7 we present the results of analysing our real-world data. The thesis ends with Chapter 8, which contains a summary and ideas for future research.

Chapter 2

Explorative data analysis

> If you torture the data long enough, it will confess.
>
> Ronald Coase

2.1 Data description

We are given two arrays of data about the NEA cod to work with in this thesis, `Ind` and `caa`. The data are provided by the Norwegian Institute of Marine Research, and are described below.

The indices `Ind`

`Ind` is a $(Y, A, 100)$-dimensional array of bootstrap samples of the indices $I_{y,a}$ described in Section 1.2 that gives us information about the abundances. These samples were made by assuming a log-normal distribution of the indices. We will not use this assumption in this thesis, and we are only going to work with the mean of these bootstrap samples. We call this mean $I_{y,a}$.

The catches `caa`

`caa` are samples from the distribution $\pi(\mathbf{C}|\mathbf{D}^C)$, where $\mathbf{C} = \{C_{y,a}\}$ is the collection of all the catches, and $\mathbf{D}^C$ are the data giving information about the catches, as described in Section 1.2. We have $M = 500$ such samples for each age-group category, and call them $C^m_{y,a}$. These were obtained through the ECA-program (Estimated Catch at Age), which was developed by Hirst et al. (2004). The age groups goes from three years old,

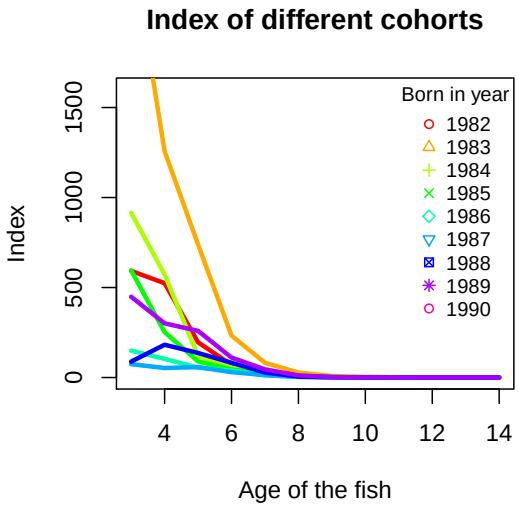
through every age (4,5,6...) until *15 years or older*. We call the last group an $A+$ group.

One thing we must emphasize about the ECA-program is that it estimates only the Norwegian catches. Norway is only responsible for about half of the catch of NEA cod (Jakobsen and Ozhigin, 2011), with Russia being the other main fishery nation. The indices on the other hand, are made by a Norwegian-Russian corporation, and should describe the entire population of NEA cod. We will describe how we handle this in Section 2.3, along with other changes we make in the data set. The original description file for the data set can be found in Appendix A.

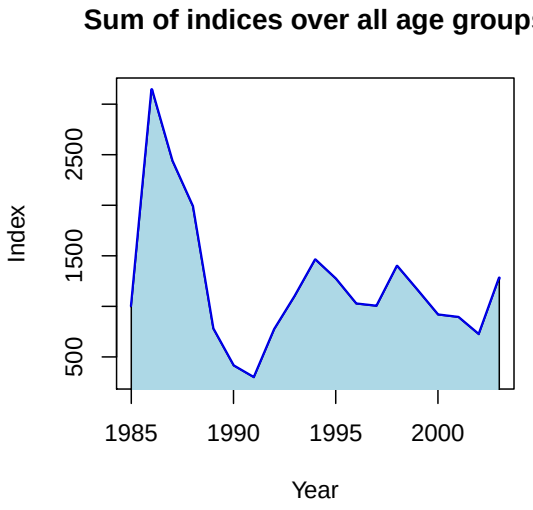
2.2 Looking at the data

The indices

Let us have a look at the indices to get an impression of how the number of fish changes over time and how the cohorts develop. We note that this first analysis is very uncertain, since we have no knowledge of the catchability parameters $\{q_a\}_{a=1}^{A}$.



(a) Displays how the different cohorts develop over time



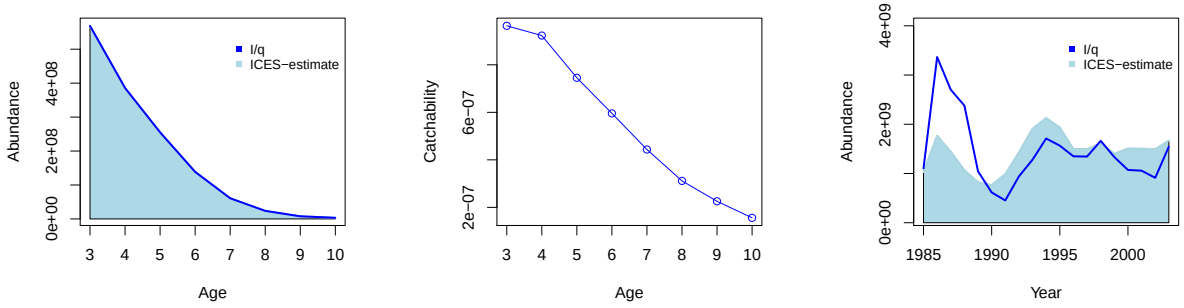
(b) Shows how the index has changed with time. Here we have summed the index of all the age groups.

Figure 2.1: The figures shows how the cohort sizes change with time, and also how the total index have changed.

Figure 2.1(a) illustrates how the cohorts develop over time. The figure displays the eight cohorts that we can follow all the way from age three to fourteen years old. We do not include the accumulation group 15+ in the plot. This is just because the A+-group contain fish from different cohorts, and thus might disturb the picture. As we can see, there are large differences in the size of the cohorts, and the largest cohort is the one with fish born in 1983. It is also interesting to look at the sum of the indices over all age groups, and get some idea about how this develops over the years. Recall from the introduction that the indices are assumed to be estimates of the abundance times some age-specific catchability. That is,

$$I_{y,a} \approx q_a N_{y,a}.$$

Because the catchability is unknown we can not directly compare the indices with the ICES estimates of the abundance. However, we note that the indices actually give a quite different picture than the ICES-estimates in Figure 1.3(b). Here we see a huge spike in the late '80s. Then we see a drop in 1990, like in the ICES-estimates, but the recovery after this is not close to being as extreme as in Figure 1.3(b). However, we can find catchabilities q_a such that the sum of $\frac{I_{y,a}}{q_a}$ over all ages resembles the plot in Figure 1.3(b). We must also emphasize that Figure 1.3(b) is made from the total estimated number of NEA cod, which also includes age groups $a = 1$ and $a = 2$.



(a) We fit the catchability such that the q_a-scaled index equals the ICES-estimates when we average over the years.

(b) Catchabilities customized to fit with the ICES estimates.

(c) Comparison of ICES-estimates and the indexes divided by the catchabilities in 2.2(b).

Figure 2.2: Catchability is estimated by requiring the mean age-estimate to equal the mean age-index divided by q_a. These catchabilities are plotted against age. With this q_a we see that there is still some difference from year to year.

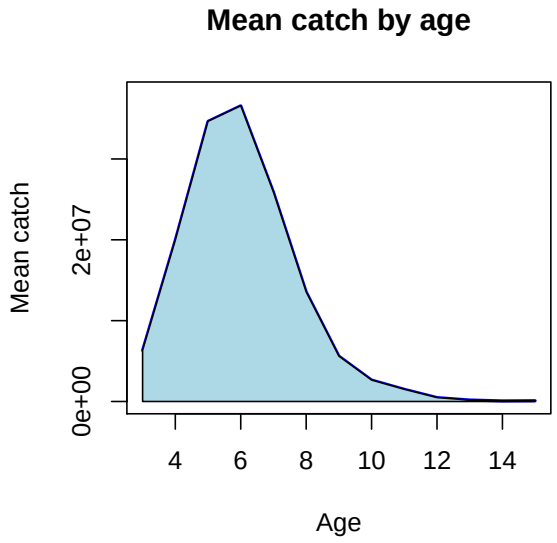
It is more interesting to consider the estimated number of fish older than three years. Then these ICES-estimates can be compared with the indices divided by some catchability. Figure 2.2(b) shows the catchability that best match the ICES-estimates. That is, we calculate

$$q_a^{ICES} = \frac{1}{Y} \sum_{y=1}^{Y} \frac{I_{y,a}}{N_{y,a}^{ICES}}.$$

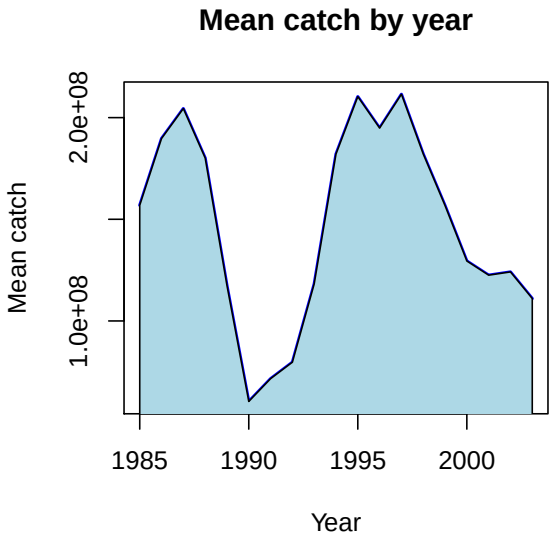
Figure 2.2 confirms that when we use q^{ICES}, the ICES-estimates equals the average of the indices divided by q^{ICES}. When we sum over all the age groups and plot against the years, the match is not that good (see Figure 2.2). The indices shows a much larger peak in the late '80s than the official estimates.

The catches

Figure 2.3(a) shows us that most of the caught NEA cod are aged between five and seven years. This is not just a coincidence. We have regulations to protect the youngest fish, because we would rather land them in a few years when they are larger. The regulations affect e.g. the grid size in the trawls. This way small fish can swim through and avoid being caught. There are few older fish caught simply because there are fewer of them.



(a) The average catch plotted against age.



(b) The average catch plotted against year. Here, we have summed over the age groups.

Figure 2.3: Shows how much fish are caught in the different age groups and in different years.

They have already been harvested, or have died by other means. Figure 2.3(b) shows how many NEA cod that are caught from year to year. The annual catch ranges from about 10 million to more than 200 million individuals in this period. Again, we recognise the drop in 1990, when many restriction were set on the fisheries.

2.3 Changes made in the data set

Both data sets are originally indexed by a season as well, corresponding to the months. For the indices the season is always $s = 2$, i.e. February. This is the month when the research surveys are done every year. The catch data was divided into all twelve months of the year. For simplicity, we only work with yearly data in this thesis. To get a hold of these, we sum the catch of every month within the year. Even though we work with yearly data we keep the season-dimension in the arrays, but set it equal to one. By doing this we can also use the computer algorithms on season-divided data (with some modifications). We are motivated to make the $A+$ group go from 10 and up, instead of 15 and up. This is primarily because there are annoying zero-observations in some of the older age classes. If we look at Figures 2.3(a) and 2.1(a) we see that the graphs flattens after the age of 10, and this justifies such a cutting. From a biological point of view it also makes sense to group these fish together. After the NEA cod i 10 years old, they are all mature [1], and they do not grow as fast as younger fish. They are predators, occupy the same area, and face the same fishery threats. In short, they are ecologically similar, and that is why we can argue biologically for a grouping. Of course, where we set the $A+$-group strongly depends on which species we consider. A 10-year old and a 15-year old individual may be fundamentally different in other species.

We then need to do something to account for the fact that we only have data about the Norwegian fisheries. One possibility is to let the natural mortality be interpreted as any cause of death apart from Norwegian fisheries, but this method has a flaw. Later on we will need to specify some prior distributions on the mortalities, and all prior research has used f (or F) as the harvest mortality for all countries together. Since we do not have access to any posterior distribution of the catch from other countries, we choose to simply scale up the Norwegian samples. The factor 2 reflects our belief about how much of the total catch that is Norwegian. Hopefully this scaling will not destroy too many of the properties of $\pi(\mathbf{C}|\mathbf{D}^C)$. The R-code used to make these changes can be found in Appendix B.

[1] This was however not the fact before 1980! (Jakobsen and Ozhigin, 2011)

Chapter 3

The Poisson-binomial model

> All models are wrong, but some are useful.
>
> George E. P. Box

In the introduction we mentioned that a virtual population analysis model is used by the Arctic Fisheries Working Group. A problem with this model is that it is deterministic. Given some starting values, all other abundances are calculated recursively. Such a model does not give us any idea about the uncertainty of the estimates. One model that do account for uncertainty, have recently been investigated by MacKean et al. (2015).

In this chapter we present a new model for the dynamics of the catch and abundance of fish. We call this model *the Poisson-binomial model*, and it will soon become apparent why. In Section 3.1 we describe and justify the model assumptions. Then, in Section 3.2 we present some theoretical results, which will be used in Section 3.3 about the model properties. Some of the model properties are merely interesting in themselves, while others will be essential when we start to build the sampling algorithms in Chapter 5. Finally, in Section 3.4 we present a model for the data $\mathbf{D}^N, \mathbf{D}^C$.

It was Geir Storvik who came up with the idea of the Poisson-binomial model to estimate fish abundance, and who first specified the assumptions that we will present in the beginning of Section 3.1. He also formulated and proved two of the theorems in Section 3.3. These theorems are marked with his name. Further work in this chapter is done by the author, with guidance from Storvik.

Throughout this chapter we will treat the model parameters $\{q_a, m_{y,a}, f_{y,a}, \lambda_{y,a}\}$ as known. That is, we will write

$$\pi(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C),$$

when we actually mean

$$\pi(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C, \{q_a, m_{y,a}, f_{y,a}, \lambda_{y,a}\}).$$

We will decide what to do with the model parameters in Chapters 4 and 5.

3.1 Model specification

Denote the abundance of fish in year y and for age group a by $N_{y,a}$. Furthermore, denote $\overline{N}_{y,a}$ to be the survivors at the end of year y, and $C_{y,a}$ to be the catch within year y. For simplicity we assume that the catch is made at the end of the year, after the natural mortality has taken place. Of course this is not true, but it has the benefit of making the math more manageable. If we view it as an approximation, it is not too unreasonable. If the time intervals are short, then the approximation is good. For now we work with yearly data, and have to accept that we use a rough approximation. In Chapter 8 we will discuss how the model can be extended to also include seasons. Written mathematically, we assume that

$$N_{y+1,a+1} = \overline{N}_{y,a} - C_{y,a}, \;\; a < A. \tag{3.1}$$

We further define A to be an $A+$ group such that

$$N_{y+1,A} = \overline{N}_{y,A-1} - C_{y,A-1} + \overline{N}_{y,A} - C_{y,A}. \tag{3.2}$$

This just means that if the fish is old, we do not care how old it is. This is convenient since we often lack data for the oldest fish, and reasonable because they are similar in many ways (see the discussion in Section 2.3). We now make the following distributional assumptions about the abundance of fish and the catch

$$N_{y,1} \sim \;\; \text{Poisson}(\lambda_{y,1}), \quad y = -A, ..., -1, 0, 1, ..., Y \tag{3.3}$$

$$\overline{N}_{y,a}|N_{y,a} \sim \text{Binom}(N_{y,a}, 1 - m_{y,a}), \tag{3.4}$$

$$C_{y,a}|\overline{N}_{y,a} \sim \text{Binom}(\overline{N}_{y,a}, f_{y,a}), \tag{3.5}$$

where $m_{y,a}$ is the probability of death by natural causes in year y for fish in age group a. Natural causes are here defined as everything apart from being caught by fisheries. $f_{y,a}$ is the probability of a fish being caught, given that it has survived all other death threats that year.

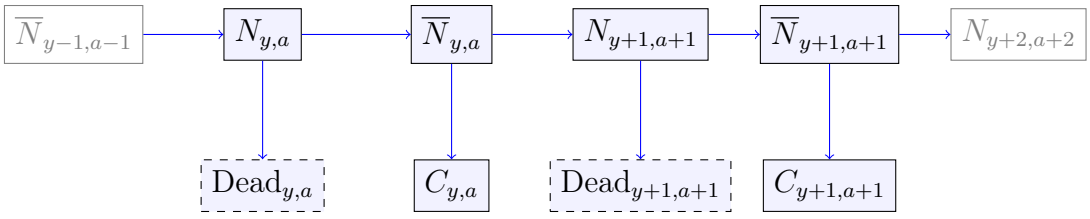


Figure 3.1: These nodes and edges form a subgraph of the dynamics of the fish abundance and catch from year to year. The variable *Dead* is not described in the dynamics, but is included here to illustrate that it is a possible state for the fish to enter.

An excerpt of the dynamics described in (3.3)-(3.5) is illustrated in Figure 3.1. Because of Equation (3.1), we immediately see that also

$$N_{y+1,a+1}|\overline{N}_{y,a} \sim \text{Binom}(\overline{N}_{y,a}, 1 - f_{y,a}) \quad a < A - 1.$$

Similarly, we see from Equation (3.2) that $N_{y+1,A}$ is the sum of the two binomially distributed variables $\overline{N}_{y,A-1} - C_{y,A-1}$ and $\overline{N}_{y,A} - C_{y,A}$. We express this as follows:

$$N_{y+1,A}|\overline{N}_{y,A}, \overline{N}_{y,A-1} \sim \text{Binom}(\overline{N}_{y,A}, 1 - f_{y,A}) + \text{Binom}(\overline{N}_{y,A-1}, 1 - f_{y,A-1}).$$

Discussion about the model assumptions

We start by arguing that the binomial assumptions in (3.4) and (3.5) are reasonable. For this purpose, we need to know what the binomial setting is:

THE BINOMIAL SETTING

1. There is a fixed number of observations n.
2. The n observations are all independent.
3. Each observation falls into one of just two categories, which for convenience we call "success" and "failure".
4. The probability of a success, call it p, is the same for each observation.

(Moore et al., 2014)

Further, we can read in Moore et al. (2014) that in the binomial setting, the count of successes X is from a binomial distribution with parameters n and p. Hence, if these four points are true, the assumptions in (3.4) and (3.5) hold. In these two equations we are given

a fixed number of observations. This number is a given number of living fish, $N_{y,a}$ and $\overline{N}_{y,a}$ respectively. Further, one of two things can happen to these fish; they can die, or continue to live. The latter alternative can surely be called a success. Thus we see that conditions 1. and 3. are fulfilled.

Conditions 2. and 4. are not that easy to justify. There is probably a location-dependence between the observations. Fish that are closer to each other are exposed to more similar conditions that affect their survival rate. There are also genetic differences between the fish that gives them different probabilities of survival. The conclusion must be that the binomial assumptions can be debated. However, we do not have a good idea about how these correlation structures are, and a model that includes them would be more complicated. For now, we live with this imperfect assumption, and leave it to future students to extend the methods of this thesis to include dependence structures. A possible way to do this is discussed in Chapter 8.

To justify the model assumption in (3.3), we need to know the *law of rare events*, or the Poisson limit theorem. This theorem is well known, and is for instance stated and proved in Billingsley (1995), page 302.

Theorem: Poisson limit theorem

Theorem 3.1.1. *If $n \to \infty$ and $p \to 0$ such that $np \to \lambda$, then*

$$\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \to \frac{\lambda^k}{k!}e^{-\lambda}.$$

Moreover, the convergence is fast and thus the Poisson distribution is an excellent approximation for large n and small p.

We can now use this theorem to argue that model assumption (3.3) is reasonable. Recall from Section 1.1 that for NEA cod, the number of eggs released in each batch ranges from 60-310 thousand eggs per female, and an egg reaches recruitment age with a probability of about 0.00065 percent. Now let (1) the millions of eggs be the *fixed number of observations* n. Furthermore, (3) define it as a success to reach recruitment age, and a failure not to. Before we know anything about where an egg is laid, it is reasonable to assume (4) that the probability of success if the same for each egg. If we are also comfortable assuming that (2)

the eggs survive independent of each other, we are in the binomial setting.

Now if the binomial assumption holds, we are only a short step away from the Poisson distribution. $p \approx 0.00065$ is certainly a small probability, and $60 - 310$ thousand times the number of spawning females is definitely a large n. Thus the Poisson limit theorem can be applied, and the Poisson assumption (3.3) is well justified.

Though we have only discussed the model assumptions with regard to NEA cod, the argumentation will be similar for other populations of fish.

3.2 Theoretical background

Before we start to investigate the properties of the Poisson-binomial model, we need to present a few theoretical results. These will be used frequently in the next section. Two of the results are about the Poisson and binomial distributions, while the remaining are about something called *blocked paths*.

The Poisson and binomial distributions

The Poisson and binomial distributions are closely related. As we just noted in Theorem 3.1.1 the Poisson distribution is the limit distribution of a binomial distribution, when the number of trials n goes to infinity and the product np stays constant. That is,

$$\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \to \frac{\lambda^k}{k!}e^{-\lambda}.$$

Since the pmf's are on similar forms it is possible to find many elegant results about the relationship between them, and two such results are presented here. These theorems will be very useful in Section 3.3.

Theorem

Theorem 3.2.1. *Assume* $X \sim Poisson(\lambda)$ *and* $Y|X \sim Binomial(X, p)$*. Then marginally*

$$Y \sim Poisson(\lambda p),$$
$$Z = X - Y \sim Poisson(\lambda(1-p)),$$

and Y *and* Z *are independent.*

The content of Theorem 3.2.1 is described in Examples 4.4.1 and 4.4.2 in *Statistical Inference* by Casella and Berger (1990).

Theorem

Theorem 3.2.2. *Assume that*

$$X_1 \sim Poisson(\lambda_1),$$
$$X_2 \sim Poisson(\lambda_2),$$

and that X_1 *and* X_2 *are independent. Then*

$$X_1|(X_1 + X_2) \sim Binomial(X_1 + X_2, \frac{\lambda_1}{\lambda_1 + \lambda_2}).$$

Theorem 3.2.2 is stated in Exercise 4.15, in *Statistical Inference* by Casella and Berger (1990), with the proof given in the solutions manual of Casella et al. (2001).

Blocked paths

We also need to know the concept of a *blocked path*. This idea comes from the theory of Directed Acyclic Graphs (DAGs). The book *Modelling and Reasoning with Bayesian Networks* by Darwiche (2009) is used as a reference here. First, let us define what a path and a valve is. These two terms are only described loosely in the text by Darwiche (2009), and not strictly defined, but the following definition should capture the essentials.

Definition: Path and valve

Definition 3.2.1. *A **path** is a trail of unique nodes and edges. The path must start and end with a node. A path that consist of three or more nodes, can be split into **valves**, which is a node on the path, and one edge going to, and one edge going from that node.*

Before we present the DAG we will, like Darwiche (2009), define

$$I_{Pr}(X, Z, Y) \tag{3.6}$$

to mean that X is conditionally independent of Y given Z under the probability measure Pr.

Darwiche (2009) does not give a formal definition of a DAG, but in Santanu (2013) (page 9) we can find the definition of a directed graph. Here we define the DAG, by mixing Santanu (2013)'s definition of directed graphs, and the notion of acyclic graphs.

Definition: DAG

Definition 3.2.2. *A **directed graph** contains a set of nodes and a set of directed edges. Furthermore, a **directed acyclic graph** (DAG) is a directed graph with no directed cycles.*

No directed cycles means that a state can never be revisited. In the theory of DAGs we use the terms neighbours, parents and children. The neighbours of a node W is any node that is directly connected to W through an edge. In Figure 3.2, $\{B, D, E\}$ are the neighbours of C. Neighbours are further divided into two classes, depending on the direction of the edge. The node the edge is leaving is called the *parent* of the node where the edge ends. The nodes that are "hit" by an edge is then called the *child* (of its parent(s)). Note that a node can be both a parent and a child. If we again look at the node C in Figure 3.2, we see that its parent is B and that its children are $\{D, E\}$.

Also, there are three different types of *valves* that can occur in a DAG. These are illustrated in Figure 3.2 and described below.

- A sequential valve ($\rightarrow W \rightarrow$) arises when W is a parent of one of its neighbours and a

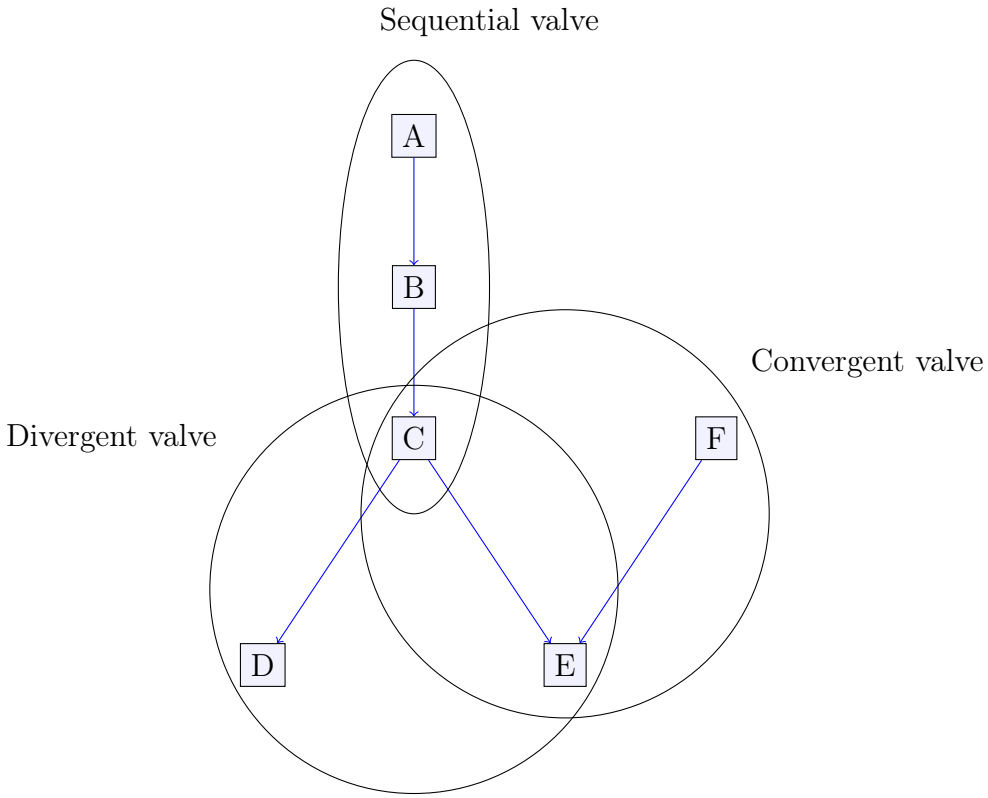


Figure 3.2: An example of a DAG. An ellipse is drawn around each of the three different valves.

child of the other.

- A divergent valve ($\leftarrow W \rightarrow$) arises when W is a parent of both neighbours.
- A convergent valve ($\rightarrow W \leftarrow$) arises when W is a child of both neighbours.

A valve can be either open or closed. Whether a valve is open or closed is important to determine independence structures in a DAG. The conditions for being closed are different for the three types of valves, and are described below. Recall the notation $I_{Pr}(X, Z, Y)$, which means that X is independent of Y if Z is known.

- A sequential valve on the path from X to Y, ($\rightarrow W \rightarrow$), is closed iff the variable W appears in Z.
- A divergent valve on the path from X to Y, ($\leftarrow W \rightarrow$), is closed iff the variable W appears in Z.
- A convergent valve on the path from X to Y, ($\rightarrow W \leftarrow$), is closed iff neither W nor any of its descendants appear in Z.

Now we have all the vocabulary we need to define the concept of d-separation.

Definition: (Darwiche, 2009)

Definition 3.2.3. *Let X, Y, and Z be disjoint sets of nodes in a DAG G. We will say that X and Y are d-separated by Z, written $dsep_G(X, Z, Y)$, iff every path between a node in X and a node in Y is blocked by Z. Here, a path is blocked by Z iff at least one valve on the path is closed given Z.*

Further, we see that d-separation actually implies independence in the probability distribution induced by a Bayesian network.

Theorem: (Darwiche, 2009)

Theorem 3.2.3. *If Pr is a probability distribution induced by a Bayesian network then*

$$I_{Pr}(X, Z, Y) \text{ if } dsep_G(X, Z, Y). \tag{3.7}$$

From the definition and theorem just presented, we can derive the following corollary.

Corollary

Corollary 3.2.1. *Let X, Y, and Z be disjoint sets of nodes in a DAG G, and assume there exists a unique path from X to Y. If any valve on this path from X to Y is closed given Z, then $I_{Pr}(X, Z, Y)$. Here Pr is the probability distribution induced by the Bayesian Network.*

Proof. This follows directly from Definition 3.2.3 and Theorem 3.2.3. □

In the next section we will use Corollary 3.2.1 frequently. When we say that a path is blocked, we always keep in mind that this implies d-separation, which in turn implies conditional independence.

3.3 Properties of the model

The model we specified in Section 3.1 has many useful properties. As we will see soon, we are able to determine the marginal distributions of $N_{y,a}$ and $C_{y,a}$. We will also be able to find the entire joint distribution of all abundances given the catches. This turns out to be a distribution we can easily draw samples from. Many results in this section are proved using the blocking argument described in the previous section. To use that argument we need to be sure that we are working with a DAG.

Theorem

Theorem 3.3.1. *Figure 3.1 on page 19 is a DAG with only sequential and divergent valves. Moreover, there is never more than one path from a node to another node.*

Proof. This can easily be seen from Figure 3.1. □

Remark Note that when we consider the last age group in the end of this section, convergent valves will appear.

The following theorem gives us the marginal distributions of $\{\overline{N}_{y,a}\}$,$\{N_{y,a}\}$ and $\{C_{y,a}\}$. A version of this result was originally stated and proved by Geir Storvik.

Theorem: (Geir Storvik)

Theorem 3.3.2. *The model specification implies that for $a < A$*

$$\overline{N}_{y,a} \sim Poisson(\lambda_{y,a}(1 - m_{y,a})), \tag{3.8}$$

$$N_{y,a} \sim Poisson(\lambda_{y,a}), \tag{3.9}$$

$$C_{y,a} \sim Poisson(\lambda_{y,a}(1 - m_{y,a})f_{y,a}), \tag{3.10}$$

where

$$\lambda_{y,a} = \lambda_{y-1,a-1}(1 - m_{y-1,a-1})(1 - f_{y-1,a-1}).$$

Further, $N_{y+1,a+1}$ and $C_{y,a}$ are independent.

Proof. Theorem 3.2.1 and model assumptions (3.3)-(3.5) directly gives

$$\overline{N}_{1,a} \sim \text{Poisson}(\lambda_{1,a}(1 - m_{1,a})),$$
$$C_{1,a} \sim \text{Poisson}(\lambda_{1,a}(1 - m_{1,a})f_{1,a}),$$
$$N_{1,a} \sim \text{Poisson}(\lambda_{1,a}(1 - m_{1,a})(1 - f_{1,a})),$$

which by induction gives the first part of the proof. Further, both $C_{1,a}$ and $N_{2,a+1}$ are derived from the Poisson distributed variable $\overline{N}_{1,a}$, and

$$C_{1,a} = N_{2,a+1} - \overline{N}_{1,a}.$$

Thus it follows from Theorem 3.2.1 that $C_{1,a}$ and $N_{2,a+1}$ are independent. □

There are two interesting results that can be proved using Theorem 3.3.2. We can show that the catch in one year is independent of *all* future abundances and catches in the cohort - not just the abundance in the next year. We can also show that the abundance in one year is independent of *all* earlier catches in the cohort - not only the catch in the previous year. These results are very interesting, but are they reasonable from a biological point of view? If not, this is a disadvantage of our model. However, from a mathematical point of view, it can be very advantageous! We start by showing that catch is independent of all future abundances and catches in the cohort.

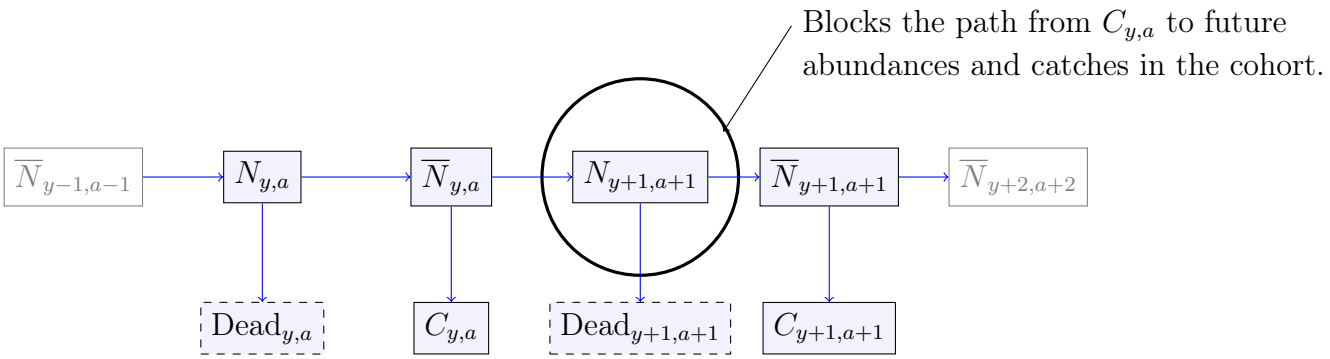


Figure 3.3: Shows how the path from $C_{y,a}$ to future abundances is blocked by $N_{y+1,a+1}$.

Theorem

Theorem 3.3.3. *The catch $C_{y,a}$ is independent of all the future abundances and catches in the cohort $N_{y+k,a+k}$ and $C_{y+k,a+k}, k \geq 1$. That is,*

$$\pi(C_{y,a}|N_{y+1,a+1}, N_{y+2,a+2}..., C_{y+1,a+1}, C_{y+2,a+2}...) = \pi(C_{y,a}).$$

Proof. We know that $C_{y,a}$ is independent of every future $N_{y+k,a+k}$ for $k \geq 2$ and $C_{y+k,a+k}$ for $k \geq 1$, if $N_{y+1,a+1}$ is given. That is because the path from $C_{y,a}$ to $\{\{N_{y+k,a+k}\}_{k\geq 2}, \{C_{y+k,a+k}\}_{k\geq 1}\}$ is blocked by $N_{y+1,a+1}$, and thus independence follows from Corollary 3.2.1. The blocking is illustrated in Figure 3.3.

We also know from Theorem 3.3.2 that $C_{y,a}$ is independent of $N_{y+1,a+1}$. Then it is true that

$$\begin{aligned}\pi(C_{y,a}|N_{y+1,a+1}, N_{y+2,a+2}, ..., C_{y+1,a+1}, C_{y+2,a+2}...) &= \pi(C_{y,a}|N_{y+1,a+1}) \\ &= \pi(C_{y,a}).\end{aligned}$$

□

Now we turn to the result that says that the abundance is not only independent of the catch of the cohort in the previous year, but in *all* previous years. This result will be very useful when we find the simultaneous distribution of all the abundances given all the catches.

Theorem

Theorem 3.3.4. *The abundance $N_{y,a}$ is independent of all the previous catches in the cohort, $C_{y-k,a-k}$ for $k \geq 1$. That is,*

$$\pi(N_{y,a}|C_{y-a+1,1}, ..., C_{y-1,a-1}) = \pi(N_{y,a}).$$

Proof. From Bayes rule we know that

$$\begin{aligned}
&\pi(N_{y,a}|C_{y-a+1,1}, ..., C_{y-1,a-1}) \\
&\propto \pi(N_{y,a}, C_{y-a+1,1}, ..., C_{y-1,a-1}) \\
&\propto \pi(C_{y-a+1,1}|C_{y-a+2,2}, ..., C_{y-1,a-1}, N_{y,a})\pi(C_{y-a+2,2}|C_{y-a+3,3}, ..., C_{y-1,a-1}, N_{y,a}) \\
&\times ... \times \pi(C_{y-2,a-2}|C_{y-1,a-1}, N_{y,a})\pi(C_{y-1,a-1}|N_{y,a})\pi(N_{y,a}).
\end{aligned}$$

Because of the independence structure we found in Theorem 3.3.2, $\pi(C_{y-1,a-1}|N_{y,a}) \propto 1$ w.r.t. $N_{y,a}$. Now let us look at the other terms. We illustrate the argument for the first term. From the rules of marginal probability distributions we know that we can write

$$\begin{aligned}
&\pi(C_{y-a+1,1}|C_{y-a+2,2}, ..., C_{y-1,a-1}, N_{y,a}) \\
&= \sum_{N_{y-a+2,2}=1}^{\infty} \pi(C_{y-a+1,1}|N_{y-a+2,2}, C_{y-a+2,2}, ..., C_{y-1,a-1}, N_{y,a}) \\
&\times \pi(N_{2,a-y+3}|C_{y-a+2,2}, ..., C_{y-1,a-1}, N_{y,a}).
\end{aligned}$$

We then observe from Figure 3.4 that $N_{y-a+2,2}$ separates $C_{y-a+1,1}$ from

$$\{C_{y-a+2,2}, ..., C_{y-1,a-1}, N_{y,a}\}.$$

Further, recall that blocking implies independence, by Corollary 3.2.1. We also recall that Theorem 3.3.2 states that $N_{y-a+2,2}$ and $C_{y-a+1,1}$ are marginally independent. Hence, it is true that

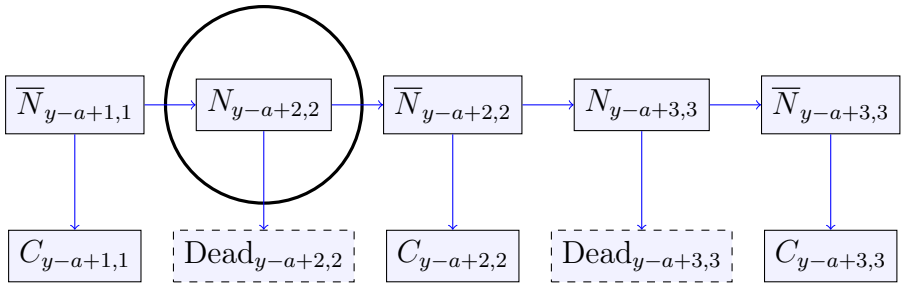


Figure 3.4: Shows how the path from $C_{y-a+1,1}$ to future abundances and catches is blocked by $N_{y-a+2,2}$.

$$
\begin{aligned}
&\pi(C_{y-a+1,1}|C_{y-a+2,2},...,C_{y-1,a-1},N_{y,a}) \\
&= \sum_{N_{y-a+2,2}=1}^{\infty} \pi(C_{y-a+1,1}|N_{y-a+2,2})\pi(N_{y-a+2,2}|C_{y-a+2,2},...,C_{y-1,a-1},N_{y,a}) \\
&= \sum_{N_{y-a+2,2}=1}^{\infty} \pi(C_{y-a+1,1})\pi(N_{y-a+2,2}|C_{y-a+2,2},...,C_{y-1,a-1},N_{y,a}) \\
&= \pi(C_{y-a+1,1}) \sum_{N_{y-a+2,2}=1}^{\infty} \pi(N_{y-a+2,2}|C_{y-a+2,2},...,C_{y-1,a-1},N_{y,a}) \\
&= \pi(C_{y-a+1,1}).
\end{aligned}
$$

To show that

$$\pi(C_{y-a+2,2}|C_{y-a+3,3},...,C_{y-1,a-1},N_{y,a}) = \pi(C_{y-a+2,2}),$$

we use the same argument, but condition on $N_{y-a+3,3}$ instead. Then we see that the argument applies to all the terms, and the only thing that is a function of $N_{y,a}$, is the marginal distribution

$$\pi(N_{y,a}).$$

□

Remark Given $\mathbf{C}$ we have the constraints

$$N_{y+1,a+1} \leq \ N_{y,a} - C_{y,a}, \quad a < A - 1.$$

These constraints can be hard to consider in forwards simulation; we might end up simulating an abundance that is smaller than the catch in the same year. This problem vanish if we

simulate backwards instead.

This motivates us to look at a backwards representation of the Poisson-binomial model. To build this theory we need to know more about the simultaneous distribution of all the abundances given all the catches. The following subsection is dedicated to this distribution.

3.3.1 Simultaneous distribution of N|C

In this subsection, we first look at the ages before the $A+$ group, or the dynamics of a cohort. The $A+$ group will then be handled separately at the end of this subsection. The main result is that we can prove that the distribution of the abundance in a cohort from year to year, given the catches, can be written as a product of Poisson pmf's. This is a very useful result that makes it possible to draw samples from this simultaneous distribution, step by step. That is, if we know the catches, we are able to sample form $\mathbf{N}|\mathbf{C}$, by sequentially sampling from

$$\pi(N_{y,a}|N_{y+1,a+1}, C_{y,a}),$$

which is actually a Poisson distribution. The reason why this is an important result will become apparent in Chapter 5. In short, it turns out that when we have samples of $\mathbf{N}|\mathbf{C}$, we can use importance sampling techniques to obtain samples from

$$\pi(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C),$$

without getting too much trouble while calculating the importance weights.

Before we can present and prove the main theorem about the simultaneous distribution of the abundances given the catches, we need some lemmas. As a natural first step, we will simplify the problem and find the simultaneous distribution of $N_{y,1}, N_{y+1,2}|C_{y,1}, C_{y+1,2}$. Moreover, before we can even do this, we need to state and prove three small lemmas.

Lemma

Lemma 3.3.1. *We have that for $a < A - 2$*

$$\pi(N_{y+a,a+1}|C_{y+a,a+1}) = p(N_{y+a,a+1} - C_{y+a,a+1}|\lambda_{y+a,a+1}(1 - f_{y+a,a+1}(1 - m_{y+a,a+1}))),$$

where $p(x|\theta)$ denotes the pmf of a Poisson distribution with parameter θ.

Proof. It immediately follows from the model assumptions (3.4) and (3.5) that

$$C_{y+a-1,a}|N_{y+a-1,a} \sim \text{Binomial}(N_{y+a-1,a}, f_{y+a-1,a}(1 - m_{y+a-1,a})),$$

and from Theorem 3.3.2 we know that marginally

$$N_{y+a-1,a} \sim \text{Poisson}(\lambda_{y+a-1,a}).$$

Then it follows from Theorem 3.2.1 that

$$\begin{aligned} C_{y+a-1,a} &\sim \text{Poisson}(\lambda_{y+a-1,a} f_{y+a-1,a}(1 - m_{y+a-1,a})), \\ N_{y+a-1,a} - C_{y+a-1,a} &\sim \text{Poisson}(\lambda_{y+a-1,a}(1 - f_{y+a-1,a}(1 - m_{y+a-1,a}))). \end{aligned}$$

In addition, $C_{y+a-1,a}$ is independent of $N_{y+a-1,a} - C_{y+a-1,a}$. Now note that we can write

$$N_{y+a-1,a} = C_{y+a-1,a} + (N_{y+a-1,a} - C_{y+a-1,a}).$$

If the catch $C_{y+a-1,a}$ is known, the only stochastic part here is the difference between the abundance and the catch. Thus the density of the abundance given the catch is given by

$$\begin{aligned} \pi(N_{y+a-1,a}|C_{y+a-1,a}) &= \pi(N_{y+a-1,a} - C_{y+a-1,a}|C_{y+a-1,a}) \\ &= \pi(N_{y+a-1,a} - C_{y+a-1,a}) \\ &= p(N_{y+a-1,a} - C_{y+a-1,a}|\lambda_{y+a-1,a}(1 - f_{y+a-1,a}(1 - m_{y+a-1,a}))). \end{aligned}$$

□

The next lemma shows a similar result, and its proof uses similar ideas.

Lemma

Lemma 3.3.2. *We have that*

$$\pi(N_{y+a-1,a}|N_{y+a,a+1}, C_{y+a-1,a}) \propto p(N_{y+a-1,a} - N_{y+a,a+1} - C_{y+a-1,a}|\lambda_{y+a-1,a} m_{y+a-1,a}),$$

for any $a < A$, where $p(x|\theta)$ denotes the pmf of a Poisson distribution with parameter θ.

Proof. When we know $N_{y+a,a+1}$ and $C_{y+a-1,a}$ we also know the sum

$$\overline{N}_{y+a-1,a} = N_{y+a,a+1} + C_{y+a-1,a}.$$

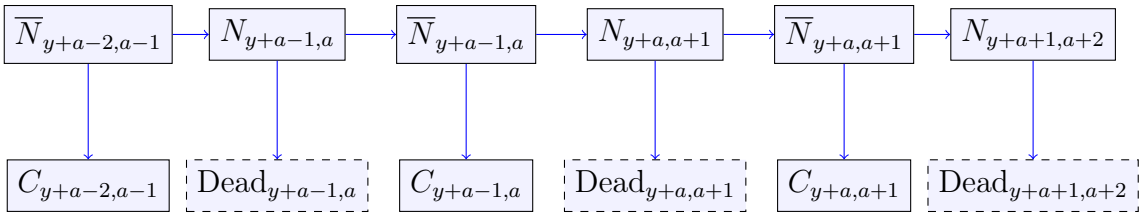


Figure 3.5: Dynamic to illustrate the blocking arguments in the proofs of Theorems 3.3.2, 3.3.3 and 3.3.6.

We see from the dynamics in Figure 3.5 that $\overline{N}_{y+a-1,a}$ blocks the path from the abundance $N_{y+a-1,a}$ to the abundance $N_{y+a,a+1}$ and the catch $C_{y+a-1,a}$. Hence it follows from Corollary 3.2.1 that

$$\pi(N_{y+a-1,a}|\overline{N}_{y+a-1,a}, N_{y+a,a+1}, C_{y+a-1,a}) = \pi(N_{y+a-1,a}|\overline{N}_{y+a-1,a}).$$

One of our model assumptions (3.4) is that

$$\overline{N}_{y+a-1,a}|N_{y+a-1,a} \sim \text{Binomial}(N_{y+a-1,a}, 1 - m_{y+a-1,a}),$$

and in Theorem 3.3.2 we proved that marginally

$$N_{y+a-1,a} \sim \text{Poisson}(\lambda_{y+a-1,a}).$$

Then it follows from Theorem 3.2.1 that

$$N_{y+a-1,a} - \overline{N}_{y+a-1,a} \sim \text{Poisson}(\lambda_{y+a-1,a} m_{y+a-1,a}),$$

and that $N_{y+a-1,a} - \overline{N}_{y+a-1,a}$ is independent of $\overline{N}_{y+a-1,a}$. Then our result follows since

$$\begin{aligned}
\pi(N_{y+a-1,a}|N_{y+a,a+1}, C_{y+a-1,a}) &= \pi((N_{y+a-1,a} - \overline{N}_{y+a-1,a}) + \overline{N}_{y+a-1,a}|\overline{N}_{y+a-1,a}) \\
&= \pi(N_{y+a-1,a} - \overline{N}_{y+a-1,a}|\overline{N}_{y+a-1,a}) \\
&= p(N_{y+a-1,a} - N_{y+a,a+1} - C_{y+a-1,a}|\lambda_{y+a-1,a} m_{y+a-1,a}).
\end{aligned}$$

□

Finally, here comes the last lemma we need before we can say something about the distribution of $N_{y,1}, N_{y+1,2}|C_{y,1}, C_{y+1,2}$.

Lemma

Lemma 3.3.3. *Given the catch $C_{y+a,a+1}$, the abundance $N_{y+a,a+1}$ is independent of the catch $C_{y+a-1,a}$ in the previous year. In other words*

$$\pi(N_{y+a,a+1}|C_{y+a-1,a}, C_{y+a,a+1}) = \pi(N_{y+a,a+1}|C_{y+a,a+1}).$$

Proof. From Bayes rule we know that

$$\pi(N_{y+a,a+1}|C_{y+a-1,a}, C_{y+a,a+1}) \propto \pi(C_{y+a-1,a+1}|N_{y+a,a+1}, C_{y+a,a+1})\pi(N_{y+a,a+1}|C_{y+a,a+1}).$$

We use a blocking argument (see Figure 3.5) to conclude that given $N_{y+a,a+1}$, $C_{y+a-1,a}$ is independent of $C_{y+a,a+1}$, such that

$$\pi(C_{y+a-1,a+1}|N_{y+a,a+1}, C_{y+a,a+1}) = \pi(C_{y+a-1,a+1}|N_{y+a,a+1}).$$

But we also have Theorem 3.3.2 saying that $N_{y+a,a+1}$ and $C_{y+a-1,a}$ are independent. Hence

$$\begin{aligned}\pi(N_{y+a,a+1}|C_{y+a-1,a}, C_{y+a,a+1}) &\propto \pi(C_{y+a-1,a})\pi(N_{y+a,a+1}|C_{y+a,a+1}) \\ &\propto \pi(N_{y+a,a+1}|C_{y+a,a+1}),\end{aligned}$$

and the result follows. □

We are now ready to show what the distribution of the abundance in the first two years of a cohort is, given the catch in these years. When we have this theorem at hand, we will actually be able to specify the entire simultaneous distribution of all abundances given all catches in a cohort. So let us have a look at it.

Theorem

Theorem 3.3.5. *We have that*

$$\begin{aligned}\pi(N_{y,1}, N_{y+1,2}|C_{y,1}, C_{y+1,2}) \propto\ & p(N_{y,1} - N_{y+1,2} - C_{y,1}|\lambda_{y,1}m_{y,1}) \\ & \times p(N_{y+1,2} - C_{y+1,2}|\lambda_{y+1,2}(1 - f_{y+1,2}(1 - m_{y+1,2}))).\end{aligned}$$

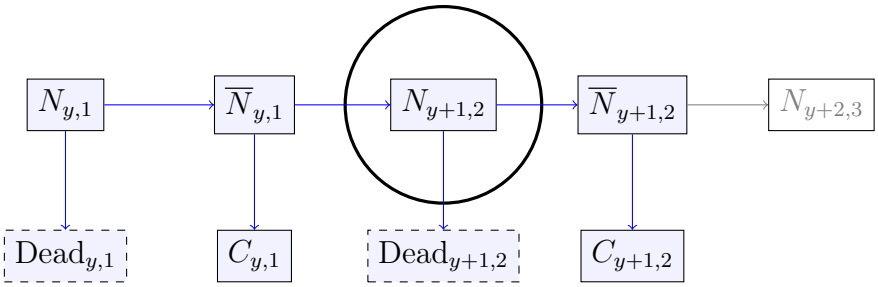


Figure 3.6: Shows how the path from $C_{y,1}$ to future abundances and catches is blocked by $N_{y+1,a}$.

Proof. First we see that

$$\pi(N_{y,1}, N_{y+1,2}|C_{y,1}, C_{y+1,2}) = \pi(N_{y,1}|N_{y+1,2}, C_{y,1}, C_{y+1,2})\pi(N_{y+1,2}|C_{y,1}, C_{y+1,2}).$$

Further, from the dynamics in Figure 3.6, we see that $N_{y+1,2}$ blocks the path from $N_{y,1}$ to $C_{y+1,2}$. Hence,

$$\pi(N_{y,1}, N_{y+1,2}|C_{y,1}, C_{y+1,2}) = \pi(N_{y,1}|N_{y+1,2}, C_{y,1})\pi(N_{y+1,2}|C_{y,1}, C_{y+1,2}).$$

As a special case of Lemma 3.3.2 we know that

$$\pi(N_{y,1}|N_{y+1,2}, C_{y,1}) = p(N_{y,1} - N_{y+1,2} - C_{y,1}|\lambda_{y,1}m_{y,1}).$$

Now let us take a closer look at the second term, $\pi(N_{y+1,2}|C_{y,1}, C_{y+1,2})$. From Lemma 3.3.3 we know that

$$\pi(N_{y+1,2}|C_{y,1}, C_{y+1,2}) = \pi(N_{y+1,2}|C_{y+1,2}).$$

which is a special case of the distribution we investigated in Lemma 3.3.1. Thus

$$\begin{aligned}\pi(N_{y,1}, N_{y+1,2}|C_{y,1}, C_{y+1,2}) \propto\ & p(N_{y,1} - N_{y+1,2} - C_{y,1}|\lambda_{y,1}m_{y,1}) \\ & \times p(N_{y+1,2} - C_{y+1,2}|\lambda_{y+1,2}(1 - f_{y+1,2}(1 - m_{y+1,2}))).\end{aligned}$$

□

We can now go even further and look at the simultaneous distribution of all the abundances given all the catches over $a < A$ years. This is the main result of this section.

Theorem: Distribution of abundances given catches in a cohort

Theorem 3.3.6. *The abundance in a cohort over $a < A$ years, conditional on the catches, has the following distribution*

$$\begin{aligned}&\pi(N_{y,1},...,N_{y+a-1,a}|C_{y,1},...,C_{y+a-1,a})\\ &\propto \pi(N_{y,1},...,N_{y+a-2,a-1}|N_{y+a-1,a},C_{y,1},...,C_{y+a-2,a-1})\pi(N_{y+a-1,a}|C_{y+a-1,a}),\end{aligned} \quad (3.11)$$

where the first part can be split up in the following way

$$\begin{aligned}&\pi(N_{y,1},...,N_{y+a-2,a-1}|N_{y+a-1,a},C_{y,1},...,C_{y+a-2,a-1})\\ &\propto \pi(N_{y,1}|N_{y+1,2},C_{y,1})\times ...\times \pi(N_{y+a-2,a-1}|N_{y+a-1,a},C_{y+a-2,a-1}),\end{aligned}$$

and each of these pmf's are on the form

$$\begin{aligned}&\pi(N_{y+a-2,a-1}|N_{y+a-1,a},C_{y+a-2,a-1})\\ &= p(N_{y+a-2,a-1}-N_{y+a-1,a}-C_{y+a-2,a-1}|\lambda_{y+a-2,a-1}m_{y+a-2,a-1}).\end{aligned}$$

The last part of equation (3.11) *is also a Poisson pmf, but on the following form*

$$\pi(N_{y+a-1,a}|C_{y+a-1,a}) = p(N_{y+a-1,a}-C_{y+a-1,a}|\lambda_{y+a-1,a}(1-f_{y+a-1,a}(1-m_{y+a-1,a}))).$$

Proof. A special case of Lemma 3.3.1 and Theorem 3.3.5 show us that this is true for $a = 1$ and $a = 2$. We will prove this is true in general by induction. Assume the claim is true for some $a' < A - 1$, and consider

$$\pi(N_{y,1},...,N_{y+a',a'+1}|C_{y,1},...,C_{y+a',a'+1}).$$

Using Bayes rule we can write this as

$$\begin{aligned}&\pi(N_{y,1},...,N_{y+a',a'+1}|C_{y,1},...,C_{y+a',a'+1})\\ &\propto \pi(N_{y,1},...,N_{y+a'-2,a'-1}|N_{y+a'-1,a'},N_{y+a',a'+1},C_{y,1},...,C_{y+a',a'+1})\\ &\times \pi(N_{y+a',a'+1},N_{y+a'-1,a'}|C_{y,1},...,C_{y+a',a'+1}).\end{aligned}$$

In Figure 3.5 we see that given $N_{y+a'-1,a'}$, the previous abundances $\{N_{y,1},...,N_{y+a'-2,a'-1}\}$ are conditionally independent of $N_{y+a',a'+1}$. So we can write

$$
\begin{aligned}
&\pi(N_{y,1},...,N_{y+a',a'+1}|C_{y,1},...,C_{y+a',a'+1}) \\
\propto\ &\pi(N_{y,1},...,N_{y+a'-2,a'-1}|N_{y+a'-1,a'},C_{y,1},...,C_{y+a'-2,a'-1}) \\
\times\ &\pi(N_{y+a',a'+1},N_{y+a'-1,a'}|C_{y,1},...,C_{y+a',a'+1}).
\end{aligned}
$$

By the induction hypothesis, $\pi(N_{y,1},...,N_{y+a'-2,a'-1}|N_{y+a'-1,a'},C_{y,1},...,C_{y+a'-2,a'-1})$ is proportional to the product

$$\pi(N_{y,1}|N_{y+1,2},C_{y,1}) \times ... \times \pi(N_{y+a'-2,a'-1}|N_{y+a'-1,a'},C_{y+a'-2,a'-1})$$

of Poisson pmf's. Further we see that

$$
\begin{aligned}
&\pi(N_{y+a',a'+1},N_{y+a'-1,a'}|C_{y,1},...,C_{y+a',a'+1}) \\
\propto\ &\pi(N_{y+a',a'+1},N_{y+a',a'+1}|C_{y+a'-1,a'},C_{y+a',a'+1}) \\
\times\ &\pi(C_{y,1},...,C_{y+a'-2,a'-1}|N_{y+a'-1,a'},N_{y+a',a'+1},C_{y+a'-1,a'},C_{y+a',a'+1}).
\end{aligned}
$$

Now take a look at Figure 3.5 again. We immediately see that

$$
\begin{aligned}
&\pi(C_{y,1},...,C_{y+a'-2,a'-1}|N_{y+a'-1,a'},N_{y+a',a'+1},C_{y+a'-1,a'},C_{y+a',a'+1}) \\
=\ &\pi(C_{y,1},...,C_{y+a'-2,a'-1}|N_{y+a'-1,a'}),
\end{aligned}
$$

because $N_{y+a'-1,a'}$ blocks the path from $\{C_{y,1},...,C_{y+a'-2,a'-1}\}$ to

$$\{N_{y+a',a'+1},C_{y+a'-1,a'},C_{y+a',a'+1}\}.$$

In Theorem 3.3.4 we showed that $N_{y+a'-1,a'}$ is marginally independent of every previous catch. Further, because of the symmetry of independence structures, it is also true that

$$
\begin{aligned}
&\pi(C_{y,1},...,C_{y+a'-2,a'-1}|N_{y+a'-1,a'}) \\
\propto\ &1,
\end{aligned}
$$

as a function of $N_{y+a'-1,a'}$. So we end up with

$$\pi(N_{y+a'-1,a'},N_{y+a',a'+1}|C_{y,1},...,C_{y+a',a'+1}) = \pi(N_{y+a'-1,a'},N_{y+a',a'+1}|C_{y+a'-1,a'},C_{y+a',a'+1}).$$

Just like in Theorem 3.3.5, we can show that this is the product

$$
\begin{aligned}
&\pi(N_{y+a'-1,a'},N_{y+a',a'+1}|C_{y+a'-1,a'},C_{y+a',a'+1}) \\
\propto\ &p(N_{y+a'-1,a'}-N_{y+1,2}-C_{y+a'-1,a'}|\lambda_{y+a'-1,a'}m_{y+a'-1,a'}) \\
\times\ &p(N_{y+a',a'+1}-C_{y+a',a'+1}|\lambda_{y+a',a'+1}(1-f_{y+a',a'+1}(1-m_{y+a',a'+1}))).
\end{aligned}
$$

Thus we see that the hypothesis holds for $a'+1$ as well, and the theorem is proved. □

The theorem we just proved says that given everything else the abundance of a cohort *only* depend on the abundance in the following year, and the catch in the same year. This enables us to sample from the abundances in a cohort, given some starting values of the abundances and all the catches. In the last part of this section we will try to sew the cohorts together in the $A+$ group.

What happens when we reach the last age group?

The results considered in the previous section concerned the case when $a < A - 1$. We will now see what happens in the last step, when we involve the $A+$ group. This is visualized in Figure 3.7.

Here we will investigate the dependence structure between the variables in the figure, as well as their distribution. In Theorem 3.3.7 we show that the abundance in the start and end of a year is Poisson distributed, also in the last age group. To prove this we first look at year $y = 2$.

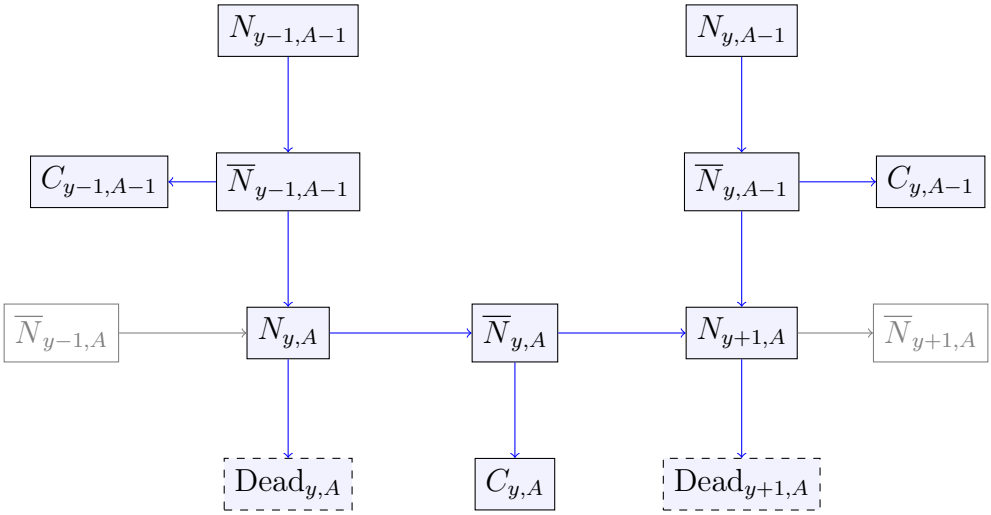


Figure 3.7: An excerpt of the dynamic of the fish abundance when the last age group is involved. This figure displays two cohorts coming "down" to the terminal age group A.

Lemma

Lemma 3.3.4.

$$\overline{N}_{1,A} \sim Poisson(\lambda_{1,A}(1 - m_{1,A}), \tag{3.12}$$

$$N_{2,A} \sim Poisson(\lambda_{2,A}), \tag{3.13}$$

$$\overline{N}_{2,A} \sim Poisson(\lambda_{2,A}(1 - m_{2,A}), \tag{3.14}$$

where

$$\lambda_{2,A} = \lambda_{1,A}(1 - f_{1,A})(1 - m_{1,A}) + \lambda_{1,A-1}(1 - f_{1,A-1})(1 - m_{1,A-1}).$$

Further, $N_{2,A}$ is independent of the catches $C_{1,A}$ and $C_{1,A-1}$. Also $\overline{N}_{1,A}$ and $\overline{N}_{1,A-1}$ are independent.

Proof. Recall from the model assumptions that

$$N_{1,A} \sim \text{Poisson}(\lambda_{1,A}),$$
$$\overline{N}_{1,A}|N_{1,A} \sim \text{Binomial}(N_{1,A}, (1 - m_{1,A})).$$

Using Theorem 3.2.1 it follows directly that, marginally

$$\overline{N}_{1,A} \sim \text{Poisson}(\lambda_{1,A}(1 - m_{1,A})).$$

In addition, we know from Theorem 3.3.2 that marginally

$$\overline{N}_{1,A-1} \sim \text{Poisson}(\lambda_{1,A-1}(1 - m_{1,A-1})).$$

We can also conclude that $\overline{N}_{1,A}$ and $\overline{N}_{1,A-1}$ are independent because they are derived from the independent variables $N_{1,A}$ and $N_{1,A-1}$ (see model assumptions).

Now the number of fish in age group A in year 2 is those of $\overline{N}_{1,A}$ and $\overline{N}_{1,A-1}$ who survive fishery. Call these survivors X and Y. Again, we use Theorem 3.2.1 to see that

$$X \sim \text{Poisson}(\lambda_{1,A}(1 - m_{1,A})(1 - f_{1,A})),$$
$$Y \sim \text{Poisson}(\lambda_{1,A-1}(1 - m_{1,A-1})(1 - f_{1,A-1})),$$

and X and Y are independent because $\overline{N}_{1,A}$ and $\overline{N}_{1,A-1}$ are so. Also, X is independent of the catch $C_{1,A} = \overline{N}_{1,A} - X$ and Y is independent of the catch $C_{1,A-1} = \overline{N}_{1,A-1} - Y$. Hence it is true that,

$$
\begin{aligned}
N_{2,A} &= X + Y \\
&\sim \text{Poisson}(\lambda_{2,A}),
\end{aligned}
$$

where

$$\lambda_{2,A} = \lambda_{1,A}(1 - f_{1,A})(1 - m_{1,A}) + \lambda_{1,A-1}(1 - f_{1,A-1})(1 - m_{1,A-1}).$$

Further, $N_{2,A}$ is independent of the two catches $C_{1,A-1}$ and $C_{1,A}$. □

Theorem

Theorem 3.3.7. *We have that*

$$N_{y,A} \sim Poisson(\lambda_{y,A}), \tag{3.15}$$

$$\overline{N}_{y,A} \sim Poisson(\lambda_{y,A}(1 - m_{y,A})), \tag{3.16}$$

where

$$\lambda_{y,A} = \lambda_{y-1,A}(1 - f_{y-1,A})(1 - m_{y-1,A}) + \lambda_{y-1,A-1}(1 - f_{y-1,A-1})(1 - m_{y-1,A-1}).$$

Proof. We know from the model assumptions and Lemma 3.3.4 that it is true for $y = 1$ and $y = 2$. The generality easily follows from induction, and by using Theorem 3.2.1. □

The following theorem has been formulated and proved a bit differently by Geir Storvik earlier. There have been some editing to match the notation used in this thesis.

Theorem: (Mostly Geir Storvik)

Theorem 3.3.8. *We have that*

$$
\begin{aligned}
&\pi(N_{y,A-1}, \overline{N}_{y,A-1}, N_{y,A} | \boldsymbol{C}_{y,*}, \boldsymbol{N}_{y+1,*}) \\
&\propto \pi(N_{y,A-1} | \overline{N}_{y,A-1}, \boldsymbol{C}_{y,A-1}) \pi(N_{y,A} | \overline{N}_{y,A-1}, \boldsymbol{C}_{y,A}, \boldsymbol{C}_{y,A-1}) \pi(\overline{N}_{y,A-1} | N_{y+1,A}, C_{y,A-1}, C_{y,A})
\end{aligned}
$$

where

$$
\begin{aligned}
&\overline{N}_{y,A-1} | N_{y+1,A}, C_{y,A-1}, C_{y,A} \\
&\sim Binom(N_{y+1,A} + C_{y,A-1} + C_{y,A}, \tfrac{\lambda_{y,A-1}(1-m_{y,A-1})}{\lambda_{y,A-1}(1-m_{y,A-1}) + \lambda_{y,A}(1-m_{y,A})}),
\end{aligned} \tag{3.17}
$$

and

$$\pi(N_{y,A-1}|\overline{N}_{y,A-1}, \boldsymbol{C}_{y,A-1}) = p(N_{y,A-1} - \overline{N}_{y,A-1}|\lambda_{y,A-1}m_{y,A-1}),$$
$$\pi(N_{y,A}|\overline{N}_{y,A-1}, \boldsymbol{C}_{y,A}, \boldsymbol{C}_{y,A-1}) = p(N_{y,A} - \overline{N}_{y,A}|\lambda_{y,A}m_{y,A}),$$

where $\overline{N}_{y,A}$ *is given by the sum of the now known variables*

$$\overline{N}_{y,A} = N_{y+1,A} + C_{y,A-1} + C_{y,A} - \overline{N}_{y,A-1}. \tag{3.18}$$

The notation $\boldsymbol{N}_{y,*}$ *denotes the set of all abundances in year* y*. That is,*

$$\boldsymbol{N}_{y,*} = \{N_{y,a}\}_{a=1}^{A}.$$

Proof. Let us first explain why the sum in (3.18) is true (keep an eye on Figure 3.7 on page 38). The abundance $N_{y+1,A}$ is a sum of the survivors from $\overline{N}_{y,A}$ and $\overline{N}_{y,A-1}$. The ones that survive are those that are not caught. So it is clear that

$$N_{y+1,A} = \overline{N}_{y,A} + \overline{N}_{y,A-1} - C_{y,A-1} - C_{y,A}.$$

A bit of basic algebra takes us from here to Equation (3.18). Now Bayes rule gives us that

$$\pi(N_{y,A-1}, \overline{N}_{y,A-1}, N_{y,A}|\mathbf{C}_{y,*}, \mathbf{N}_{y+1,*})$$
$$\propto \pi(N_{y,A-1}|\overline{N}_{y,A-1}, N_{y,A}, \mathbf{C}_{y,*}, \mathbf{N}_{y+1,*})\pi(N_{y,A}|\overline{N}_{y,A-1}, \mathbf{C}_{y,*}, \mathbf{N}_{y+1,*})\pi(\overline{N}_{y,A-1}|\mathbf{C}_{y,*}, \mathbf{N}_{y+1,*}).$$

For each of these three distributions we use Figure 3.7 to find the nodes that blocks the path to all the nodes in the future. This gives us that

$$\pi(N_{y,A-1}, \overline{N}_{y,A-1}, N_{y,A}|\mathbf{C}_{y,*}, \mathbf{N}_{y+1,*})$$
$$\propto \pi(N_{y,A-1}|\overline{N}_{y,A-1}, C_{y,A-1})\pi(N_{y,A}|\overline{N}_{y,A-1}, C_{y,A}, C_{y,A-1})\pi(\overline{N}_{y,A-1}|N_{y+1,A}, C_{y,A-1}, C_{y,A}).$$

Further, we show which distributions these are. To prove the assertion in Equation (3.17), we will utilize Theorem 3.2.2. Now given $C_{y,A-1}, C_{y,A}$ and $N_{y+1,A}$ we know the sum

$$\overline{N}_{y,A-1} + \overline{N}_{y,A} = N_{y+1,A} + C_{y,A-1} + C_{y,A}.$$

We know from the previous theorem that this is the sum of two independent random variables from Poisson distributions with parameters

$$\lambda_{y,A-1}(1 - m_{y,A-1}),$$
$$\lambda_{y,A}(1 - m_{y,A}).$$

Theorem 3.2.2 then implies that

$$\begin{aligned}
&\overline{N}_{y,A-1}|C_{y,A-1}, C_{y,A}, N_{y+1,A} \\
&\sim \text{Binom}(\overline{N}_{y,A-1} + \overline{N}_{y,A}, \tfrac{\lambda_{y,A-1}(1-m_{y,A-1})}{\lambda_{y,A-1}(1-m_{y,A-1})+\lambda_{y,A}(1-m_{y,A})}) \\
&= \text{Binom}(N_{y+1,A} + C_{y,A-1} + C_{y,A}, \tfrac{\lambda_{y,A-1}(1-m_{y,A-1})}{\lambda_{y,A-1}(1-m_{y,A-1})+\lambda_{y,A}(1-m_{y,A})}).
\end{aligned}$$

Finally, by arguments like in the proof of Lemma 3.3.2, we see that

$$\begin{aligned}
\pi(N_{y,A-1}|\overline{N}_{y,A-1}, C_{y,A-1}) &= p(N_{y,A-1} - \overline{N}_{y,A-1}|\lambda_{y,A-1}m_{y,A-1}), \\
\pi(N_{y,A}|\overline{N}_{y,A-1}, C_{y,A}, C_{y,A-1}) &= p(N_{y,A} - \overline{N}_{y,A}|\lambda_{y,A}m_{y,A}).
\end{aligned}$$

□

Remark The proof we just gave is not complete, as we have only considered the distribution when future abundances and catches are given. When we looked at Figure 3.7, we were only concerned about which nodes blocked the path to future nodes. Still we strongly believe that Theorem 3.3.8 holds in general. In order to prove this, it should be sufficient to prove an equivalent of Theorem 3.3.4 for the last age group. This proof is not included mainly because it would take a lot of time, and partly because the the arguments would be a repetition of earlier arguments.

Summary of the model properties

Through this section we have learned both interesting and useful properties of the Poisson-binomial model. We have found the marginal distributions of all the abundances and catches. We saw that in this model, the abundance in one year-age group is independent of all the previous catches, if nothing else is given. Similarly, we could flip the coin and say that the catch in one year-age group is independent of all the future catches and abundances. After some more work, we arrived at the main result of this section. Namely that we can split the simultaneous distribution of all abundances, given all the catches, up in smaller parts (combine Theorems 3.3.6 and 3.3.8). These parts are all Poisson or binomial distributions that we will be able to draw samples from recursively.

3.4 Model for the data $\mathbf{D}^N, \mathbf{D}^C$

We have available catch data $\mathbf{D}^C$, and some indices $\mathbf{D}^N = \{I_{y,a}\}$ that gives some information about the abundances $\mathbf{N}$. The two data sets $\mathbf{D}^N$ and $\mathbf{D}^C$ are assumed to be conditionally independent, given the latent states $\mathbf{N}, \mathbf{C}$.

The catch data has a likelihood $\pi(\mathbf{D}^C|\mathbf{C})$ which is (very) difficult to evaluate. We will however assume that we have samples $\mathbf{C}^b, b = 1, ..., B$ from the posterior distribution

$$\pi(\mathbf{C}|\mathbf{D}) \propto \tilde{\pi}(\mathbf{C})\pi(\mathbf{D}^{\mathbf{C}}|\mathbf{C}),$$

where $\pi(\mathbf{D}^C|\mathbf{C})$ is the true likelihood for the data, but where $\tilde{\pi}(\mathbf{C})$ is some working prior. Hence, we will not go further and specify any of the distributions involving $\mathbf{D}^C$.

Binomial indices

The indices are based on trawl observations, where a number of fish are caught by trawling specific regions. We assume that the indices come from a binomial distribution with parameters $(N_{y,a}, q_a)$. This assumption is easy to work with in our framework, where everything follows Poisson or binomial distributions. One benefit is that we can now find the marginal distribution of the indices.

Corollary

Corollary 3.4.1. *Assume that* $I_{y,a}|N_{y,a}, q_a \sim Binom(N_{y,a}, q_a)$. *Then marginally*

$$I_{y,a} \sim Poisson(\lambda_{y,a}q_a). \tag{3.19}$$

Proof. In Theorem 3.3.2 we showed that $N_{y,a} \sim \text{Poisson}(\lambda_{y,a})$. We have also assumed that $I_{y,a}|N_{y,a} \sim \text{Binomial}(N_{y,a}, q_a)$. Then it follows from Theorem 3.2.1 that marginally

$$I_{y,a} \sim \text{Poisson}(\lambda_{y,a}q_a).$$

□

Figure 3.8 shows which nodes the data $\{\mathbf{D}^N, \mathbf{D}^C\}$ gives information about.

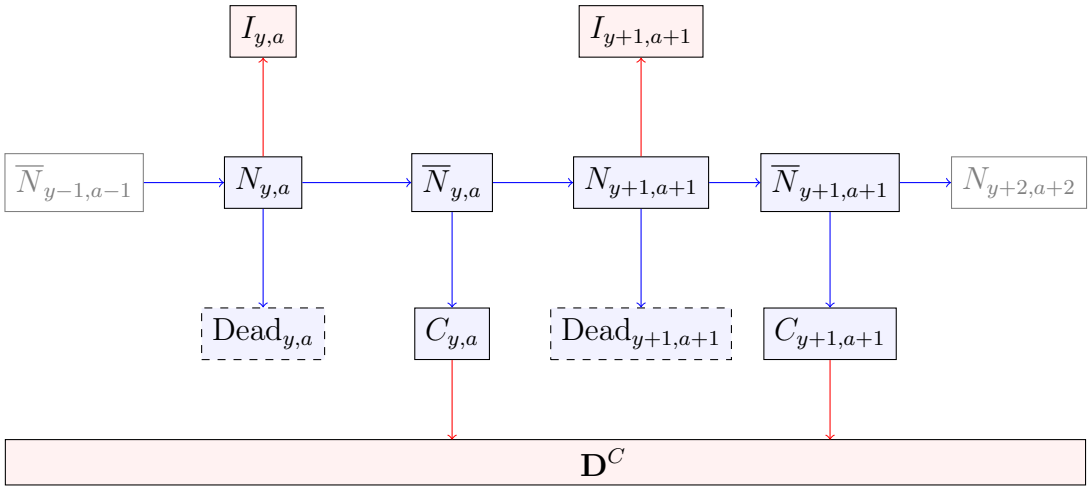


Figure 3.8: The blue boxes and lines is an excerpt of the dynamics of the fish abundance and catch from year to year. The red boxes shows the data sources.

Chapter 4

Identifiability and estimability

> Mathematics is the art of giving the same name to different things.
>
> Henri Poincaré

In Chapter 3 we saw how the simultaneous distribution $\pi(\mathbf{N}|\mathbf{C})$ can be split up in Poisson and binomial pmf's. To draw samples from $\pi(\mathbf{N}|\mathbf{C})$ is an important step on the way to obtaining samples from $\pi(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C)$, as we will see in Chapter 5. However, in Chapter 3 we always treated the model parameters $\{\lambda_{y,s,a}\}$, $\{m_{y,s,a}\}$, $\{f_{y,s,a}\}$ and $\{q_a\}$ as known. These parameters are of course not known, and we need to go through a bit of trouble to estimate them. Exactly how we estimate the parameters will be the theme of the next chapter. This chapter is more theoretical, and we aim to show under which conditions the Poisson-binomial model is *identifiable*.

We will start this chapter by defining what identifiability is, and how it is related to estimability. Then, in the second and last section we investigate the identifiability problems in the Poisson-binomial model.

4.1 Introduction to identifiability and estimability

The definitions and results we present in this section can also be found in the review article *On identifiability of parametric statistical models* by Paulino and Pereira (1994). Before we define what identifiability is, we need to introduce some concepts and notation. In general we consider a statistical model $\{\mathscr{Y}, \mathscr{A}, \mathscr{P}\}$ where $\mathscr{Y}$ is the sample space, $\mathscr{A}$ is a σ-algebra defined on the sample space, and $\mathscr{P}$ is a family of probability measures on $(\mathscr{Y}, \mathscr{A})$, defined by $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$, where Θ is the parametric space.

Definition: (Paulino and Pereira, 1994)

Definition 4.1.1. *Two points of Θ, θ_0 and θ_1 are said to be* ***observationally equivalent****, and we write $\theta_0 \sim \theta_1$, if*

$$P_{\theta_0}(A) = P_{\theta_1}(A),$$

for all $A \in \mathscr{A}$.

The following corollary is easily derived from this definition.

Corollary

Corollary 4.1.1. *$\sim$ is an equivalence relation.*

In Paulino and Pereira (1994) it is simply stated that it is "easy to see" that $\sim$ is an equivalence relation. Still, a proof is included here to erase any doubt.

Proof. $\sim$ is an equivalence relation because it satisfies the three properties of an equivalence relation:

Reflexivity

Because P_{θ_0} and P_{θ_0} is the exact same probability measure, it is obvious that $P_{\theta_0}(A) = P_{\theta_0}(A)$ for all $A \in \mathscr{A}$.

Symmetry

This follows directly from the properties of the equality sign.

$$P_{\theta_1}(A) = P_{\theta_2}(A) \implies P_{\theta_2}(A) = P_{\theta_1}(A).$$

Transitivity

Assume that $P_{\theta_0} = P_{\theta_1}$ and that $P_{\theta_1} = P_{\theta_2}$. Then, again from the properties of the equality sign, it follows that $P_{\theta_0} = P_{\theta_2}$.

□

Because $\sim$ is an equivalence relation, we can make a partition in Θ of the equivalence classes $[\theta_0] = \{\theta_1 \in \Theta : P_{\theta_1} = P_{\theta_0}\}$. This partition is called the quotient set of Θ according to the equivalence relation $\sim$, and we denote it by $\Theta/\sim$. We are now ready to define the concept of identifiability.

Definition: (Paulino and Pereira, 1994)

Definition 4.1.2. *Θ (or the model $\boldsymbol{P}$) is said to be identifiable if $[\theta] = \{\theta\}, \forall\theta \in \Theta$, i.e., if $\Theta/\sim$ is the finest partition, $\Theta/\sim = \{\{\theta\}, \theta \in \Theta\}$.*

In simpler language, this means that a model is identifiable if

$$P_{\theta_1}(A) = P_{\theta_2}(A) \Leftrightarrow \theta_1 = \theta_2.$$

We also need to know the concept of an identifiable function.

Definition: (Paulino and Pereira, 1994)

Definition 4.1.3. *A function $\phi(\theta)$ is said to be identifiable if the corresponding induced partition is identifiable, i.e. $\forall\theta_0, \theta_1 \in \Theta$, $\theta_1 \in [\theta_0] \implies \phi(\theta_1) = \phi(\theta_0)$.*

This just means that if ϕ is identifiable, the probability measure is the same if the function ϕ of the model parameters is the same. That is

$$P_{\theta_1}(A) = P_{\theta_2}(A) \implies \phi(\theta_1) = \phi(\theta_2).$$

An identifiable parameter is actually just a special case of an identifiable function. To see this, note that if we let $\phi(\theta) = \theta$, Definitions 4.1.2 and 4.1.3 means exactly the same. We will now define what an estimable function is, and say something about how identifiability is related to estimability.

Definition: (Paulino and Pereira, 1994)

Definition 4.1.4. *A function $\phi(\theta)$ is estimable if it admits an unbiased estimator.*

Other definitions of estimability are also common (Paulino and Pereira, 1994). Some define it in a weaker sense, and only require that the function $\phi(\theta)$ admits a *consistent* estimator. Others again use the term to describe something called *uniqueness of estimation*. We will not discuss the different terms further, but work with the definition of Paulino and Pereira (1994), which is the most common - according to them. As we see from the following theorem, estimability implies identifiability.

Theorem: (Paulino and Pereira, 1994)

Theorem 4.1.1. *If $\phi(\theta)$ is an estimable function, then $\phi(\theta)$ is identifiable.*

Finally, we also need the theorem from Paulino and Pereira (1994) saying that a function of an identifiable function is identifiable.

Theorem: (Paulino and Pereira, 1994)

Theorem 4.1.2. *If ϕ is an identifiable function and ψ is a function of ϕ, then ψ is also identifiable.*

Paulino and Pereira (1994) do not include a proof for this theorem, but we prove the theorem here to make sure that it is true.

Proof. If $\phi(\theta_1) = \phi(\theta_0)$, then $\psi(\phi(\theta_1)) = \psi(\phi(\theta_0))$. Now since $\phi(\theta)$ is identifiable, we know that $\forall\theta_0, \theta_1 \in \Theta$, $\theta_1 \in [\theta_0] \implies \phi(\theta_1) = \phi(\theta_0)$. Now, if we add the implication

$$\phi(\theta_1) = \phi(\theta_0) \implies \psi(\phi(\theta_1)) = \psi(\phi(\theta_0)),$$

we see that $\psi(\phi(\theta))$ also fits in the definition of an identifiable function. □

In the next section we will use the definitions and results presented here, to investigate the identifiability problems in the Poisson-binomial model.

4.2 Identifiability and estimability of the Poisson-binomial model

Our model is over-parametrized in the sense that the parameters are functions of each other. We recall from one of our first results, Theorem 3.3.2, that

$$\lambda_{y+1,a+1} = \lambda_{y,a}(1 - m_{y,a})(1 - f_{y,a}).$$

From this relationship we see that we can write $m_{y,a}$ as a function of the λs and $f_{y,a}$. This function is given by

$$m_{y,a}(f_{y,a}, \lambda_{y+1,a+1}, \lambda_{y,a}) = 1 - \frac{\lambda_{y+1,a+1}}{\lambda_{y,a}(1 - f_{y,a})}.$$

Similarly, $f_{y,a}$ can be written as the function

$$f_{y,a}(m_{y,a}, \lambda_{y+1,a+1}, \lambda_{y,a}) = 1 - \frac{\lambda_{y+1,a+1}}{\lambda_{y,a}(1 - m_{y,a})}.$$

Since the natural mortalities and fishery mortalities are functions of one another, we only need to find one of them, together with the λs, in order to find the other ($\mathbf{m}$ or $\mathbf{f}$). This reduces our actual parameter space from $\Theta = \{\lambda, \mathbf{m}, \mathbf{f}, \mathbf{q}\}$ to $\Theta = \{\lambda, \mathbf{f}, \mathbf{q}\}$ or $\Theta = \{\lambda, \mathbf{m}, \mathbf{q}\}$.

Before we go deeper into our model, we need to present and prove a lemma about the Poisson distribution.

Lemma

Lemma 4.2.1. *In a Poisson distribution, the parameter θ is identifiable.*

Proof. The pmf of the Poisson distribution with parameter θ is given by

$$p(x|\theta) = \frac{\theta^x}{x!}e^{-\theta},\ x \in \{0, 1, 2, ...\}.$$

Assume $p(x|\theta_1) = p(x|\theta_2)\ \forall x \in \{0, 1, 2, ...\}$, then for any such pair $\theta_1, \theta_2 \in \Theta$,

$$\begin{aligned} \frac{\theta_1^x}{x!}e^{-\theta_1} &= \frac{\theta_2^x}{x!}e^{-\theta_2}, \\ e^{x\log(\theta_1)-\theta_1} &= e^{x\log(\theta_2)-\theta_2}, \\ x\log(\theta_1) - \theta_1 &= x\log(\theta_2) - \theta_2. \end{aligned}$$

Because $x = 0$ is in the support, we see that $\theta_1 = \theta_2$. Hence $[\theta] = \{\theta\}\ \forall \theta \in \Theta$, and this parameter is identifiable. □

Actually, the parameter θ in the Poisson distribution is not only identifiable, but also estimable.

Lemma

Lemma 4.2.2. *In a Poisson distribution, the parameter θ is estimable.*

Proof. It is well known that the mean is an unbiased estimator of θ, and it is also quick to prove it

$$\begin{aligned} E\big[\frac{1}{n}\sum_{i=1}^{n} X_i\big] &= \frac{1}{n}\sum_{i=1}^{n} E(X_i), \\ &= \frac{1}{n}\sum_{i=1}^{n} \theta, \\ &= \theta. \end{aligned}$$

Then it follows from Definition 4.1.4 that θ is estimable. □

It would have been sufficient just to prove that θ is estimable, because estimability implies identifiability (Theorem 4.1.1). However, the proof of identifiability gives some intuitive understanding of what identifiability is. The fact that the Poisson rate parameter is estimable, leads us to find two estimable functions in our model.

Theorem

Theorem 4.2.1. *The Poisson-binomial model has two sets of estimable, and thus identifiable, functions. These are given by*

$$\begin{aligned} \phi_{1,y,a}(\theta) &= q_a \lambda_{y,a}, \\ \phi_{2,y,a}(\theta) &= \lambda_{y,a} f_{y,a}(1 - m_{y,a}), \end{aligned}$$

for every possible pair of (y, a).

Proof. We recall from Theorem 3.3.2 that marginally

$$\begin{aligned} I_{y,a} &\sim \text{Poisson}(q_a \lambda_{y,a}), \\ C_{y,a} &\sim \text{Poisson}(\lambda_{y,a} f_{y,a}(1 - m_{y,a})). \end{aligned}$$

Then the result follows directly from the previous lemma, Lemma 4.2.2. □

From Theorem 4.1.2 it follows that any function of ϕ_1 and ϕ_2 is also identifiable. There may exist even more identifiable functions, but we have not been able to prove whether any such function exists - or does not exist. Additional identifiable functions might arise from looking at the simultaneous distributions of the indices and catches. However, we will just continue to work with the identifiable functions $\phi_{1,y,a}$ and $\phi_{2,y,a}$, and functions of them, $\{\psi(\phi_{1,y,a}, \phi_{2,y,a})\}$. This leads us to the following theorem.

Theorem

Theorem 4.2.2. *All model parameters are identifiable functions, if at least one of* $\{q_{a+1}\}$, $\{f_{y,a}\}$ *or* $\{\lambda_{y,a}\}$ *is known. We refer to equations (1)-(4) below. If* $\{q_{a+1}\}$ *is known, then the functions (2)-(4) are identifiable. If* $\{f_{y,a}\}$ *is known, then the functions (1),(3) and (4) are identifiable. Finally, if* $\{\lambda_{y,a}\}$ *is known, then the functions (1), (2), and (4) are identifiable.*

$$q_{a+1} = \frac{\phi_{1,y+1,a+1}}{\phi_{2,y,a}(\frac{1}{f_{y,a}} - 1)} = \frac{\phi_{1,y+1,a+1}}{\lambda_{y+1,a+1}}, \tag{1}$$

$$f_{y,a} = \frac{q_{a+1}\phi_{2,y,a}}{\phi_{1,y+1,a+1} + q_{a+1}\phi_{2,y,a}}, \tag{2}$$

$$\lambda_{y,a} = \frac{\phi_{1,y,a}}{q_a}, \tag{3}$$

$$m_{y,a} = 1 - \frac{q_a(q_{a+1}\phi_{2,y,a} + \phi_{1,y+1,a+1})}{q_{a+1}\phi_{1,y,a}}. \tag{4}$$

Moreover, all parameters are not identifiable only given $\{m_{y,a}\}$.

Proof. This proof contains a lot of simple algebra, to obtain the four equations in the theorem. However, the main idea is that Theorem 4.1.2 tells us that functions of identifiable functions are also identifiable. If the four equations in the theorem are true, it is easy to see that $\{q_{a+1}\}$, $\{f_{y,a}\}$, $\{\lambda_{y,a}\}$ and $\{m_{y,a}\}$ are identifiable provided one of $\{q_{a+1}\}$, $\{f_{y,a}\}$ or $\{\lambda_{y,a}\}$ is known.

However, it is not sufficient to assume $\{m_{y,a}\}$ known. That is because every natural mortality $m_{y,a}$ is a function of two other model parameters, in addition to the identifiable functions ϕ_1 and ϕ_2. When $m_{y,a}$ is *not* a one-to-one function of any of the other parameters,

assuming it known will not alone help us to identify anything else.

So, let us do the algebra to prove the four equations in the theorem are true. We need to recall from Theorems 4.2.1 and 3.3.2 that

$$\begin{aligned}
\phi_{1,y,a}(\theta) &= q_a \lambda_{y,a}, \\
\phi_{2,y,a}(\theta) &= \lambda_{y,a} f_{y,a}(1 - m_{y,a}), \\
\lambda_{y+1,a+1} &= \lambda_{y,a}(1 - m_{y,a})(1 - f_{y,a}).
\end{aligned}$$

Then we first see that

$$\begin{aligned}
\frac{\phi_{1,y+1,a+1}}{\phi_{2,y,a}(\frac{1}{f_{y,a}} - 1)} &= \frac{q_{a+1}\lambda_{y+1,a+1}}{\lambda_{y,a}(1 - m_{y,a}) f_{y,a}(\frac{1}{f_{y,a}} - 1)} \\
&= \frac{q_{a+1}\lambda_{y,a}(1 - m_{y,a})(1 - f_{y,a})}{\lambda_{y,a}(1 - m_{y,a})(1 - f_{y,a})} \\
&= q_{a+1}.
\end{aligned}$$

Next we see that

$$\begin{aligned}
\frac{q_{a+1}\phi_{2,y,a}}{\phi_{1,y+1,a+1} + q_{a+1}\phi_{2,y,a}} &= \frac{q_{a+1}\lambda_{y,a} f_{y,a}(1 - m_{y,a})}{q_{a+1}\lambda_{y+1,a+1} + q_{a+1}\lambda_{y,a} f_{y,a}(1 - m_{y,a})} \\
&= \frac{\lambda_{y,a} f_{y,a}(1 - m_{y,a})}{\lambda_{y,a}(1 - f_{y,a})(1 - m_{y,a}) + \lambda_{y,a} f_{y,a}(1 - m_{y,a})} \\
&= \frac{\lambda_{y,a} f_{y,a}(1 - m_{y,a})}{\lambda_{y,a}(1 - m_{y,a})} \\
&= f_{y,a}.
\end{aligned}$$

Then we move to the third equation and see that

$$\begin{aligned}
\frac{\phi_{1,y,a}}{q_a} &= \frac{q_a \lambda_{y,a}}{q_a}, \\
&= \lambda_{y,a}.
\end{aligned}$$

Finally, we see that

$$\begin{aligned}
1 - \frac{q_a(q_{a+1}\phi_{2,y,a} + \phi_{1,y+1,a+1})}{q_{a+1}\phi_{1,y,a}} &= 1 - \frac{q_a(q_{a+1}\phi_{2,y,a} + \phi_{1,y+1,a+1})}{q_{a+1}q_a\lambda_{y,a}} \\
&= 1 - \frac{q_{a+1}\lambda_{y,a}(1 - m_{y,a}) f_{y,a} + q_{a+1}\lambda_{1,y+1,a+1}}{q_{a+1}\lambda_{y,a}} \\
&= 1 - \frac{\lambda_{y,a}(1 - m_{y,a}) f_{y,a} + \lambda_{y,a}(1 - m_{y,a})(1 - f_{y,a})}{\lambda_{y,a}} \\
&= 1 - \frac{\lambda_{y,a}(1 - m_{y,a})}{\lambda_{y,a}} \\
&= m_{y,a}.
\end{aligned}$$

□

The question we are left with now is: Which variable should we assume known? The λs sounds like a bad choice, as they are the mean of the abundances, which is what we are mainly interested in.

There are no obvious reasons why we should choose either the catchabilities or fishery mortalities known. However, we choose to treat the catchabilities as known. These are easier to specify than the fishery mortalities, because we can assume that they are the same from year to year. This assumption would be less reasonable for the fishery mortalities, and we would then have to fine-tune even more parameters ($Y \cdot A$ instead of A). In the upcoming chapter we will present methods on how to actually estimate the model parameters, together with the idea of importance sampling.

Chapter 5

Inference methods

> An ounce of algebra is worth of a ton of verbal argument.
>
> John Maynard Smith

Our main goal is to obtain samples from the joint distribution $\pi(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C)$. In this chapter we develop two methods to achieve this. The two methods differ in the way they treat the model parameters $\{m_{y,a}\}$,$\{f_{y,a}\}$,$\{q_a\}$ and $\{\lambda_{y,a}\}$. In the first method, which we refer to as Method 1, we use the generalized method of moments (GMM) to estimate the model parameters.

In the second method, Method 2, we use an Empirical Bayes approach. Here we specify a prior distribution on each of the model parameters.

Whether we estimate the parameters directly or put prior distributions on them, we will use importance sampling (IS) to obtain samples from $\pi(\mathbf{N}, \mathbf{C}|\mathbf{D}^N, \mathbf{D}^C)$.

In this chapter we focus on presenting the theory of GMM, prior distributions and importance sampling. We will also show how these methods are applied to the Poisson-binomial model, and describe the computer algorithms we have developed. In Chapter 6 we will test the algorithms on a simulated data set, before we apply them to the real data in Chapter 7.

5.1 Method 1: Generalized method of moment estimation (GMM)

The generalized method of moments, abbreviated by GMM, is based on *moment conditions*. In general, let Y denote the data, and let θ denote the model parameters. Then we need a function g such that

$$m(\theta_0) \equiv E[g(Y_t, \theta_0)] = 0.$$

Here, θ_0 is the true parameter values. The idea of the method of moments is to replace the expectation with the sample average. That is,

$$\hat{m}(\theta) = \frac{1}{T}\sum_{t=1}^{T} g(Y_t, \theta).$$

Then we minimize the Euclidean norm of this expression with respect to θ. Formally, Ziegler (2011) define the GMM estimator as follows.

Definition: (Ziegler, 2011)

Definition 5.1.1. *A GMM estimator is any value $\hat{\theta}$ minimizing*

$$\Big(\frac{1}{T}\sum_{t=1}^{T} g(Y_t, \theta)\Big)^T \Big(\frac{1}{T}\sum_{t=1}^{T} g(Y_t, \theta)\Big).$$

The *generalization* lies in the fact that $g(Y_t, \theta)$ can be multi-dimensional. If $g(Y_t, \theta)$ is in $\mathbb{R}$, we recognise this as the usual method of moments. The GMM estimator has many desirable properties, and three of them are presented in the following theorem.

Theorem: (Ziegler, 2011)

Theorem 5.1.1. *Under some required regularity conditions, which are referred to in Ziegler (2011), it is true that*

1. *There exists a GMM estimator $\hat{\theta}$ for θ_0.*
2. *The GMM estimator $\hat{\theta}$ converges almost surely to the true parameter vector θ_0.*
3. *The GMM estimator $\hat{\theta}$ can be obtained by solving the first-order conditions*

$$\frac{1}{T}\sum_{t=1}^{T} g(Y_t, \hat{\theta}) = 0.$$

The data we have available, the catches and indices, are both marginally Poisson distributed. Hence the following result will be useful.

Lemma

Lemma 5.1.1. *For a Poisson distributed variable Y with parameter θ, the function $g(Y,\theta) = Y - \theta$ is such that*

$$E\Big[g(Y,\theta)\Big] = 0.$$

Hence, it yields a GMM estimator, and this is given by

$$\hat{\theta} = \frac{1}{T}\sum_{t=1}^{T} Y_t.$$

Because we are in one dimension, this is the same as the usual method of moments estimator.

Proof. We see that

$$\begin{aligned} E\Big[g(Y,\theta)\Big] &= E\big[Y-\theta\big] \\ &= E\big[Y\big] - \theta \\ &= 0. \end{aligned}$$

Further, the estimator $\hat{\theta} = \frac{1}{T}\sum_{t=1}^{T} Y_t$ is such that

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} g(Y_t,\hat{\theta}) &= \frac{1}{T}\sum_{t=1}^{T}\Big(Y_t - \frac{1}{T}\sum_{t=1}^{T} Y_t\Big) \\ &= 0. \end{aligned}$$

This certainly minimizes $\Big[\frac{1}{T}\sum_{t=1}^{T} g(Y_t,\hat{\theta})\Big]^2$. □

With this result in mind we start looking for GMM estimators in our model. In the previous chapter on identifiability we showed that ϕ_1 and ϕ_2 are estimable functions. Now we show how we can actually estimate them by using the GMM theory.

Theorem

Theorem 5.1.2. *Define*

$$\theta = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix},$$

where ϕ_1 is the vector of $\phi_{1,y,a} = q_a \lambda_{y,a}$ for all pairs of (y, a) in some order, and similarly ϕ_2 is the vector of all $\phi_{2,y,a} = \lambda_{y,a}(1-m_{y,a})f_{y,a}$. Then a GMM estimator is with elements

$$\hat{\phi}_{1,y,a} = I_{y,a},$$
$$\hat{\phi}_{2,y,a} = \frac{1}{B}\sum_{b=1}^{B} C_{y,a}^b,$$

for every pair of (y, a).

Proof. It follows from point 3. in Theorem 5.1.1 that we can find the GMM estimator by solving the first-order equations. The solutions of the first order equations follows from Lemma 5.1.1 because,

$$I_{y,a} \sim \text{Poisson}(\phi_{1,y,a}),$$
$$C_{y,a} \sim \text{Poisson}(\phi_{2,y,a}).$$

□

With estimates of ϕ_1 and ϕ_2, we can calculate estimates of all model parameters $\{\lambda, \mathbf{m}, \mathbf{f}, \mathbf{q}\}$ if the intensities, fishery mortalities or catchabilities are given. To be able to do this, we must recall the equations we found in Theorem 4.2.2.

$$f_{y,a} = \frac{q_{a+1}\phi_{2,y,a}}{\phi_{1,y+1,a+1} + q_{a+1}\phi_{2,y,a}},$$
$$q_{a+1} = \frac{\phi_{1,y+1,a+1}}{\phi_{2,y,a}(\frac{1}{f_{y,a}} - 1)},$$
$$\lambda_{y,a} = \frac{\phi_{1,y,a}}{q_a},$$
$$m_{y,a} = 1 - \frac{q_a(q_{a+1}\phi_{2,y,a} + \phi_{1,y+1,a+1})}{q_{a+1}\phi_{1,y,a}}.$$

To illustrate that the estimators are in fact reasonable, we will spend some time on them. We will investigate all the three cases: When $\{q_a\}$ is known, when $\{f_{y,a}\}$ is known and when $\{\lambda_{y,a}\}$ is known.

Assume $\{q_a\}$ is known

Observe that we can write

$$\begin{aligned}\hat{f}_{y,a} &= \frac{q_{a+1}\hat{\phi}_{2,y,a}}{\hat{\phi}_{1,y+1,a+1} + q_{a+1}\hat{\phi}_{2,y,a}} \\ &= \frac{\hat{\phi}_{2,y,a}}{\frac{\hat{\phi}_{1,y+1,a+1}}{q_{a+1}} + \hat{\phi}_{2,y,a}} \\ &= \frac{C_{y,a}}{\frac{I_{y+1,a+1}}{q_{a+1}} + C_{y,a}}.\end{aligned}$$

Now, $\frac{I_{y+1,a+1}}{q_{a+1}} \approx N_{y+1,a+1}$ by assumption. Recall also that $\overline{N}_{y,a} = N_{y+1,a+1} + C_{y,a}$. Thus the estimator of the fishery mortality is actually similar to the intuitive estimator

$$\hat{f}_{y,a} \approx \frac{C_{y,a}}{\overline{N}_{y,a}}.$$

That is, the fishery mortality is estimated by the ratio of the fish caught, to those that can be caught. In a similar way, we can show that the estimator

$$\hat{m}_{y,a} = 1 - \frac{q_a(q_{a+1}\hat{\phi}_{2,y,a} + \hat{\phi}_{1,y+1,a+1})}{q_{a+1}\hat{\phi}_{1,y,a}}$$

is reasonable. Observe first that we can write

$$\hat{m}_{y,a} = 1 - \frac{\hat{\phi}_{2,y,a} + \frac{\hat{\phi}_{1,y+1,a+1}}{q_{a+1}}}{\frac{\hat{\phi}_{1,y,a}}{q_a}}.$$

Again we use that $\frac{\hat{\phi}_{1,y,a}}{q_a} \approx N_{y,a}$, which gives us that

$$\begin{aligned}\hat{m}_{y,a} &\approx 1 - \frac{C_{y,a} + N_{y+1,a+1}}{N_{y,a}} \\ &= \frac{N_{y,a} - \overline{N}_{y+1,a+1}}{N_{y,a}}.\end{aligned}$$

That is, we estimate the natural mortality by the the number that is alive at the beginning of the year minus the number alive at the end of the year, and divide it by the number alive in the first place. That is also the most intuitive thing to do.

It is also easy to see that the λ is estimated in a natural way, as

$$\begin{aligned}\hat{\lambda} &= \frac{\hat{\phi}_{1,y,a}}{q_a} \\ &= \frac{I_{y,a}}{q_a}.\end{aligned}$$

Algorithm 5.1.1 displays how the parameter estimation is carried out when $\{q_a\}$ is assumed known.

Algorithm 5.1.1 GMM-estimation with known catchability

1: **for** $y \in \{1, ..., Y\}, a \in \{1, ..., A\}$ **do**
2: Let $\hat{\lambda}_{y,a} = \frac{I_{y,a}}{q_a}$.
3: Let $\frac{1}{B}\sum_b C^b_{y,a} = \bar{C}^{\bullet}_{y,a}$.
4: **for** $y \in \{1, ..., Y-1\}, a \in \{1, ..., A-1\}$ **do**
5: Let $Nbar = \lambda_{y+1,a+1} + \bar{C}^{\bullet}_{y,a}$.
6: Let $f_{y,a} = \bar{C}^{\bullet}_{y,a}/Nbar$.
7: Let $m_{y,a} = 1 - \bar{C}^{\bullet}_{y,a}/(\lambda_{y,a} f_{y,a})$.
8: Set m and f in the last age group and the last year equal to the parameters in the next to last age/year.

Assume $\{f_{y,a}\}$ is known

Now the estimator of the catchability is

$$\begin{aligned}\hat{q}_{a+1} &= \frac{\hat{\phi}_{1,y+1,a+1}}{\hat{\phi}_{2,y,a}(\frac{1}{f_{y,a}} - 1)} \\ &= \frac{I_{y+1,a+1}}{C_{y,a}(\frac{1}{f_{y,a}} - 1)} \\ &= \frac{I_{y+1,a+1}}{\frac{C_{y,a}}{f_{y,a}} - C_{y,a}}.\end{aligned}$$

The fraction $\frac{C_{y,a}}{f_{y,a}}$ in the denominator is approximately equal to $\overline{N}_{y,a}$. This follows from a simple inversion of

$$C_{y,a} \approx \overline{N}_{y,a} f_{y,a}.$$

This means that the denominator actually is an estimate of $\overline{N}_{y,a} - C_{y,a} = N_{y+1,a+1}$. Recall that the indices are assumed to be estimates of $q_{a+1} N_{y+1,a+1}$, and thus

$$\hat{q}_{a+1} \approx \frac{q_{a+1} N_{y+1,a+1}}{N_{y+1,a+1}},$$

which makes a lot of sense. When the estimator for the catchability is reasonable, it follows from the previous paragraph that the estimators of the natural mortality and the intensity is reasonable as well.

Assume $\{\lambda_{y,a}\}$ is known

If the intensities $\{\lambda_{y,a}\}$ are known, it is clear that

$$\begin{aligned} \hat{q}_a &= \frac{\hat{\phi}_{1,y,a}}{\lambda_{y,a}} \\ &= \frac{I_{y,a}}{\lambda_{y,a}}. \end{aligned} \tag{5.1}$$

is a reasonable estimator because

$$\begin{aligned} I_{y,a} &\approx q_a N_{y,a}, \\ \lambda_{y,a} &\approx N_{y,a}. \end{aligned}$$

Again, every other estimator is intuitive, now that q_a is known.

Remark Note that the previous paragraph is not entirely correct because we keep the catchability constant over the years. Hence there are Y estimators for q_a, according to Equation (5.1). The natural thing here is to average over the years, such that

$$\hat{q}_a = \frac{1}{Y} \sum_{1}^{Y} \frac{I_{y,a}}{\lambda_{y,a}}.$$

Remark We only develop computer algorithms that handle the special case when the catchability $\{q_a\}$ is known.

5.2 Method 2: Priors on model parameters

The R-code used in this section can be found in Appendix C.

In a Bayesian setting we view the model parameters as stochastic variables instead of fixed values. These variables have distributions, which we call prior distributions before the data are taken into account. Prior is a Latin word that means former or elder, and that is exactly what the prior distribution is; it represent our belief about the parameters *prior to* the observations. We are free to choose whichever prior we like, but it should represent which values we think the parameters have. Also, some priors make computation easier than others. To cite John Tsitsiklis[1]

> Choosing a prior requires the certain skill of not making a silly choice.

The first non-silly thing we can do is to use *conjugate priors*. A prior $\pi(\theta)$ is a conjugate prior for the likelihood $\pi(y|\theta)$, if the posterior distribution $\pi(\theta|y)$ is from the same family of probability distributions. Our model parameters $\{m_{y,a}, f_{y,a}, q_a\}$ are all parameters in binomial distributions. Because of this, the following result is useful to know.

Lemma: (Carlin and Louis, 2009)

Lemma 5.2.1. *The family of beta distributions is the unique conjugate prior for the binomial success probability p.*

Because of this, we will use beta priors on these parameters. That is, we specify

$$\begin{aligned} m_{y,a} &\sim Beta(\alpha_{m,y,a}, \beta_{m,y,a}), \\ f_{y,a} &\sim Beta(\alpha_{f,y,a}, \beta_{f,y,a}), \\ q_a &\sim Beta(\alpha_{q,a}, \beta_{q,a}). \end{aligned}$$

We also assume that the parameters $\{\mathbf{m}, \mathbf{f}, \mathbf{q}\}$ are independent of one another. The next task is to specify the hyperparameters, the αs and βs, such that the prior distributions represent our beliefs. This can be done in many different ways. The non-informative way would be to put a uniform prior distribution on all the parameters, which is equal to a beta distribution

[1] From MIT Open Course Ware, lecture 21: Bayesian Statistical Inference I

with parameters $\alpha = \beta = 1$. However, it is often hard to construct efficient algorithms with non-informative prior distributions. Especially when there are many parameters, which is the case here. Therefore, we make a bit more informative prior distributions.

When we turn to look for prior information, we could use an *Empirical Bayes* approach and base the prior distributions on our estimates from Section 5.1. Another approach is to put our beliefs in the estimates in the ICES (2014)-report. We will use this second approach. In all the priors we specify, we let the ICES-estimate be the mean. In fact, this is also an Empirical Bayes approach, because the AFWG have used the same data as us to obtain these estimates.

Now let us see how we specify the priors for the four classes of model parameters one by one.

The natural mortality m

As mentioned in Chapter 1, the AFWG use $M = 0.2$ for all years and age groups. Recall that this M is related to our mortality parameter in the following way

$$m = 1 - e^{-M}.$$

That is, $M = 0.2$ corresponds to $m = 1 - e^{-0.2} \approx 0.18$. Thus it is reasonable for us to choose a prior distribution that puts much weight around 0.18. A formal way to set the prior is to require that the mean of the beta distribution to equal the average ICES estimate. In general, the mean in a beta distribution with parameters (α_m, β_m) is given by

$$E(M) = \frac{\alpha_m}{\alpha_m + \beta_m}.$$

Since we want $E(M) = 0.18$, we must require that

$$\frac{\alpha_m}{\alpha_m + \beta_m} = 0.18,$$

which in turn implies

$$\beta_m = \frac{(1 - 0.18)\alpha_m}{0.18}.$$

This way we can try different values of α and let β be automatically updated to match our demand to let $E(M) = 0.18$. Note that β is now an *increasing* function of α. In addition to decide where the mean should be, we need to have an opinion about the variance in the distribution. In general, for beta distributions

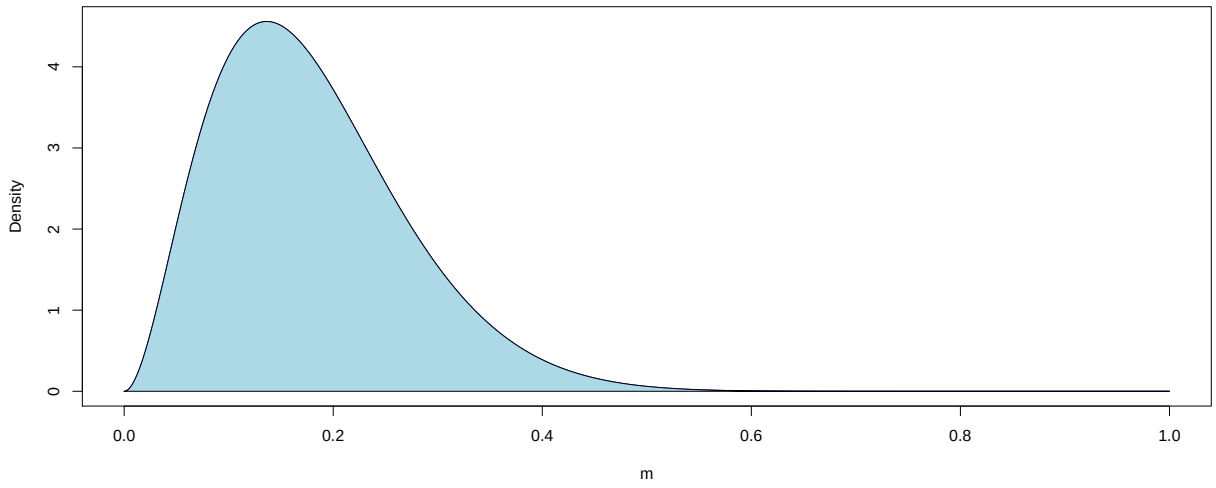


Figure 5.1: Beta distribution with $\alpha = 3$ and $\beta = 13.67$.

$$Var(M) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

It is apparent that when the values of α and β increase, the variance decrease.

After having tried different values of α, it seems graphically reasonable to choose $\alpha = 3$ and $\beta = 13.67$. The beta distribution with these parameters is shown in Figure 5.1. We use the same prior distribution for all ages and years.

The fishery mortality f

Here we will again use the ICES (2014)-report. The harvest control rule is described in this report, and a part of it says that

> ... If the TAC [total annual quota], by following such a rule, corresponds to a fishing mortality (F) lower than 0.30 the TAC should be increased to a level corresponding to a fishing mortality of 0.30.

The fisheries always fill their quota. If we also assume that the ICES estimates are good, we can assume that the instantaneous fishing mortality (F) is above 0.3. Some more prior information is that in 2013 the fishery mortality (F) was estimated to be 0.34. However, there are large differences between the age groups. The average estimated fishery mortalities in the different age groups from 1946 to 2013 are shown in Table 5.1. The numbers are collected from Table 3.19 in ICES (2014). Again we need to remember that we use a different

Table 5.1: Fishery mortalities in the different age groups. The first line in the ICES-estimates, and in the second line we have transformed them to the mortality parameter we use; f.

	3	4	5	6	7	8	9	10	11	12	+gr
F	0.005	0.035	0.1	0.176	0.296	0.366	0.412	0.362	0.321	0.264	-
f	0.005	0.034	0.095	0.16	0.26	0.31	0.34	0.30	-	-	-

parametrization than the AFWG. Like for the natural mortalities, the relationship is given by

$$f = 1 - e^{-F}.$$

The ICES-estimates transformed to our f's are also shown in Table 5.1.

In our data set we have set the $A+$ group when the fish are 10 years or older. We will use the average of the last three groups in the table, as a reference for the mortality in our 10+ group.

$$E(f_{y,10}) = 1 - e^{-\frac{0.362+0.321+0.264}{3}} = 0.30.$$

We use the same method as for the natural mortalities m. That is, we require that the mean of the beta distribution to equal the average ICES estimate. Considering e.g. age group $a = 7$, we require

$$\frac{\alpha_{f,y,7}}{\alpha_{f,y,7} + \beta_{f,y,7}} = 0.26,$$

which implies

$$\beta_{f,y,7} = \frac{(1 - 0.26)\alpha_{f,y,7}}{0.26}.$$

Again, when we now start to look for suitable parameters, we only need to adjust α. Figure 5.2 shows some suggested beta distributions. The hyperparameters we end up using are given in Table 5.2.

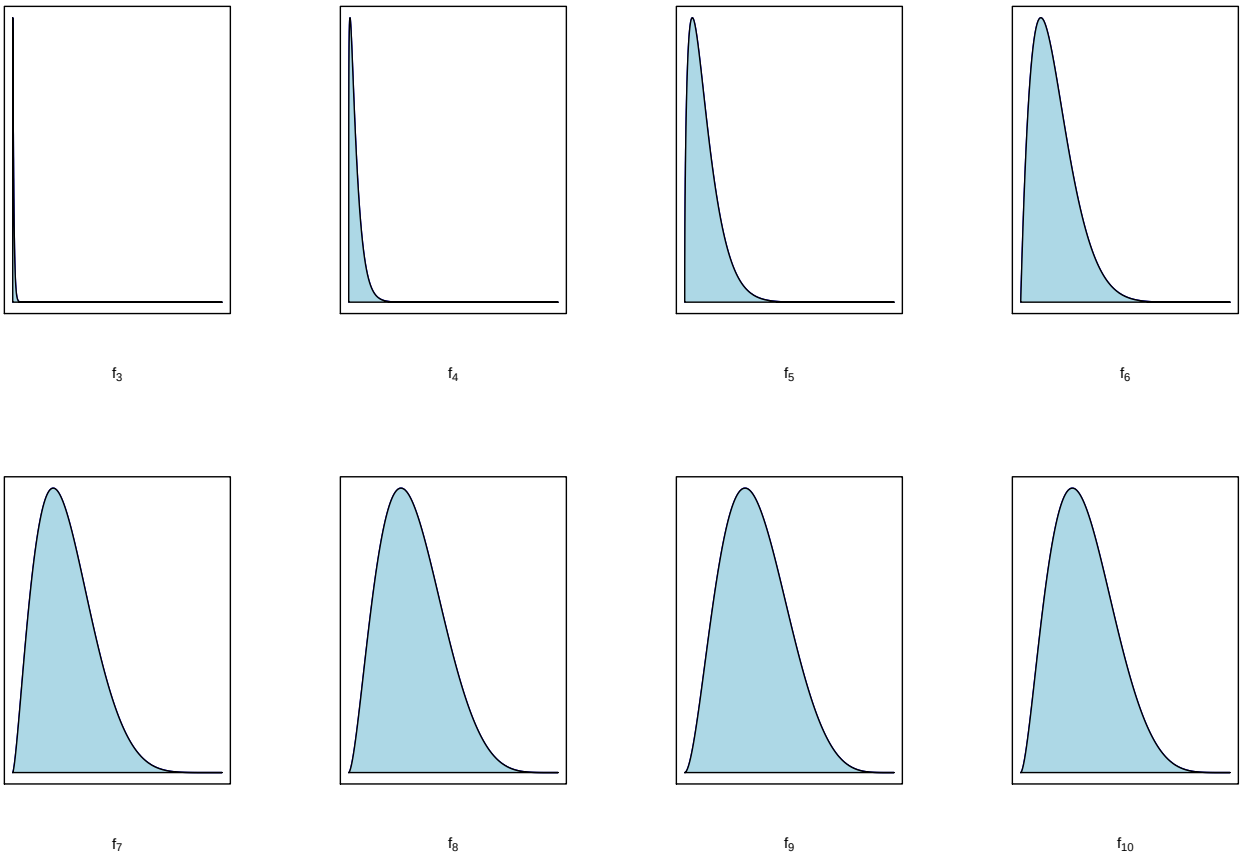


Figure 5.2: Priors on the fishery mortalities for every age group.

Table 5.2: Hyperparameters in the beta-prior on fishery mortalities $\{f_{y,a}\}$ after fine-tuning.

	3	4	5	6	7	8	9	10 +
α	1.01	1.2	1.5	2	2.5	2.7	2.9	2.7
β	201.5	33.7	14.3	10.4	7.3	6.1	5.7	6.2

The intensities λ_1

We also need to set a prior on $\lambda_1 = \{\lambda_{1,a}, \lambda_{y,1}\}_{y=1,a=2}^{Y,A}$. We recall that the λs are parameters in the following Poisson distribution,

$$N_{y,a}|\lambda_{y,a} \sim \text{Poisson}(\lambda_{y,a}).$$

Once again we wish to use a conjugate prior, and we look to the *Distributional catalog* in Carlin and Louis (2009). There we find that

Lemma: (Carlin and Louis, 2009)

Lemma 5.2.2. *The family of gamma distributions is the unique conjugate prior for the Poisson rate parameter θ.*

Based on this we will use the prior

$$\lambda \sim Gamma(\alpha, \beta),$$

with some values for α and β. The ICES (2014)-report will again provide prior information, which we can use to specify the hyperparameters. We will put the mean of our priors on the ICES estimates, but we will let them be quite flat. Tables 5.3 and 5.4 contain the information we need from Table 3.16 in the ICES (2014)-report.

Table 5.3: Estimates in year 1985.

$E(\Lambda_{y,a})$	a=3	a=4	a=5	a=6	a=7	a=8	a=9	a=10+
y=1985	528732	323311	97995	47269	20823	6481	3181	969

When we choose the hyperparameters α and β we keep in mind that for a gamma distributed variable $\lambda_{y,a}$

$$E(\lambda_{y,a}) = \alpha\beta,$$
$$\mathsf{Var}(\lambda_{y,a}) = \alpha\beta^2,$$

where β is the rate parameter, which is also called the inverse scale parameter. We will require the product $\alpha\beta$ to equal the ICES estimates, and can control the variance (a larger β, and thus smaller α, gives larger variance). We start by considering the first year, which is 1985, and the first age group. After some trials it seems reasonable to let

$$\beta_{1985,3} = 5 \cdot 10^7,$$

Figure 5.3 shows this prior distribution for $\lambda_{1985,3}$.

Further, we let

$$\beta_{1985,a} = 5 \cdot 10^7, \quad a \in \{4, 5\},$$
$$\beta_{1985,a} = 5 \cdot 10^6, \quad a \in \{6, 7\},$$
$$\beta_{1985,a} = 5 \cdot 10^5, \, a \in \{8, 9, 10+\}.$$

Plots of these distributions are not included. We also need to specify the prior distribution of the recruits every year. Just like with the parameters m and f, we will not assume any difference between the years. Hence we put the same prior on $\lambda_{y,a=3}$ for every year $y \in \{1986, ..., 2003\}$. We will set the mean of the prior equal to the average of the ICES-estimates in Table 5.4, that is

$$E(\lambda_{y,3}) = 502 \cdot 10^6.$$

Table 5.4: Estmates for $a = 3$.

$E(\Lambda_{y,a})$	1986	1987	1988	1989	1990	1991	1992	1993	1994
a=3	104737	288868	206585	174629	245930	416500	729566	904988	792583
$E(\Lambda_{y,a})$	1995	1996	1997	1998	1999	2000	2001	2002	2003
a=3	623260	444981	724332	853363	553859	616361	524294	457237	703241

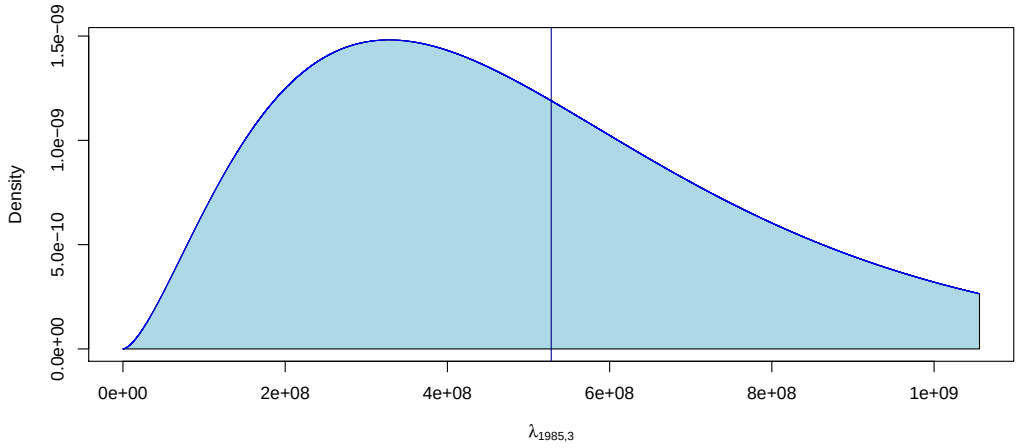


Figure 5.3: Prior for $\Lambda_{1985,3}$. The ICES estimate is at $5.28 \cdot 10^8$.

Using the same method as before, we end up choosing the hyperparameters

$$\begin{aligned} \alpha_{y,a=3} &= 2.51, \\ \beta_{y,a=3} &= 2 \cdot 10^8. \end{aligned}$$

The catchability q

We recall that in Chapter 4, we proved that all other model parameters could be estimated if the catchabilities are fixed. Without any good idea about how to estimate the catchability, our ad hoc solution is to let

$$\widehat{q}_a^{(b)} = \frac{\frac{1}{Y}\sum_{y=1}^{Y} I_{y,a}}{\lambda_{y,a}^{(b)}},$$

after we have obtained B samples of $\lambda_{y,a}$ from its prior distribution. When $\{q_a\}$ is estimated in this way, the algorithm runs quite smoothly. However, now we do not really use the information in the indices that much. The way we estimate the catchability insures us that the indices fit well with the λ's, which again are based on previous estimates.

The algorithm

The algorithm below, Algorithm 5.2.1, summarizes how we use the material of this section. All the hyperparameters in the algorithm refer to those described in this section. For each parameter we obtain B samples. These are indexed by a superscript b.

Algorithm 5.2.1 Priors on model parameters

for $b \in 1 : B$ **do**
  for $a \in 1 : A$ **do**
    for $y \in 1 : Y$ **do**
      Draw $m^b_{y,a} \sim \text{Beta}(\alpha_{m,y,a}, \beta_{m,y,a})$.
      Draw $f^b_{y,a} \sim \text{Beta}(\alpha_{f,y,a}, \beta_{f,y,a})$.
      Draw $\lambda^b_{y,a} \sim \text{Gamma}(\alpha_{\lambda,y,a}, \beta_{\lambda,y,a})$.
      Set $q^b_a = \frac{\bar{I}_{\bullet,a}}{\lambda^b_{y,a}}$

5.3 Importance sampling

Our aim is to obtain samples from the joint posterior distribution $\pi(\mathbf{N}, \mathbf{C}|D^N, D^C)$. Once we have the model parameters, obtained from either Method 1 or Method 2, we can get samples from $\pi(\mathbf{N}, \mathbf{C}|D^N, D^C)$ by using *importance sampling (IS)*.

In this section we will present the general idea of importance sampling, how we can apply it to our model, and an extension of importance sampling called *sequential importance sampling with resampling (SISR)*.

5.3.1 Introduction to importance sampling (IS)

We use Creal (2012) as a reference for this section. The general idea of importance sampling is to utilize that for any function $h(X)$

$$\begin{aligned} E^\pi[h(X)|D] &= \int h(X)\pi(X|D)dX \\ &= \int h(X)\pi(X|D)\frac{q(X|D)}{q(X|D)}dX \\ &= E^q[h(X)\frac{\pi(X|D)}{q(X|D)}] \\ &\approx \frac{1}{B}\sum_1^B h(X^b)\frac{\pi(X^b|D)}{q(X^b|D)} \\ &= \frac{1}{B}\sum_1^B h(X^b)w(X^b), \end{aligned}$$

where $w(X^b) = \frac{\pi(X^b|D)}{q(X^b|D)}$ are called the *importance weights*. This approximation is true for any proposal function $q(X^b|D)$ that is non-zero wherever $\pi(X^b|D)$ is non-zero (Creal (2012)). The computer algorithm we use to find the importance samples is a special case of Algorithm

5.3.1, which we will describe later, in Section 5.3.2. When all the X^b are drawn from the proposal distribution we can find the mean by letting $h(X) = X$. We can also choose a function $h(X)$ such that we find sample quantiles, as we will discover in Section 5.3.2. A shortcut could always be to resample the X^b's with probabilities equal to the normalised importance weights. These resampled X^b's can then be viewed as samples from the true distribution.

Effective Sample Size (ESS)

We want to avoid the situation where a few observations are given much weight, while the other weights are close to zero. Then we effectively have only a few observation. This happens if the proposal distribution is too different from the true distribution. One way to measure the effective sample size (ESS) is to look at the measure

$$\text{ESS} = \frac{1}{\sum_{b=1}^{B} w_b^2}. \tag{5.2}$$

If $w_b = \frac{1}{B}$ for all b, we see that ESS $= B$, and in the other extreme case where

$$w_b = \begin{cases} 1 & b = b^*, \\ 0 & \text{otherwise}, \end{cases} \tag{5.3}$$

for a specific b^*, we see that $ESS = 1$. Hence the interpretation of ESS is that if the number is close to B, then most observations contribute to the estimators, but if the ESS is close to one, this is not the case. There exist computational techniques that are designed to avoid that the ESS gets too low. One of them is *sequential importance sampling with resampling* (SISR), which we describe in Section 5.3.2. But first we look at how importance sampling can be applied in the Poisson-binomial model.

Importance sampling for the Poisson-binomial model

We can use importance sampling to get samples from the posterior distribution

$$\pi(\mathbf{N}, \mathbf{C}, \mathbf{m}, \mathbf{f}, \lambda, \mathbf{q} | \mathbf{D}^N, \mathbf{D}^C).$$

Under Method 1 we just use fixed estimates of the model parameters. The "posterior", like the "prior", of the parameters can in that setting be interpreted as the pmf with mass 1 on the parameter estimate. Under Method 2 the priors are those specified in Section 5.2, and

we do not know yet what the posteriors will look like, and this posterior is of great interest. To ease notation, denote the collection of the parameters by

$$\theta = \{\mathbf{m}, \mathbf{f}, \lambda, \mathbf{q}\}.$$

The importance sampling idea can now be used in the following way

$$\begin{aligned} E^{\pi}[h(\mathbf{N},\mathbf{C},\theta)|\mathbf{D}^N,\mathbf{D}^C] &= \int\int\int h(\mathbf{N},\mathbf{C},\theta)\pi(\mathbf{N},\mathbf{C},\theta|\mathbf{D}^N,\mathbf{D}^C)d\mathbf{N}d\mathbf{C}d\theta \\ &= \int\int\int h(\mathbf{N},\mathbf{C},\theta)\pi(\mathbf{N},\mathbf{C},\theta|\mathbf{D}^N,\mathbf{D}^C)\frac{q(\mathbf{N},\mathbf{C},\theta|\mathbf{D}^N,\mathbf{D}^C)}{q(\mathbf{N},\mathbf{C},\theta|\mathbf{D}^N,\mathbf{D}^C)}d\mathbf{N}d\mathbf{C}d\theta \\ &= E^{q}[h(\mathbf{N},\mathbf{C},\theta)\frac{\pi(\mathbf{N},\mathbf{C},\theta|\mathbf{D}^N,\mathbf{D}^C)}{q(\mathbf{N},\mathbf{C},\theta)|\mathbf{D}^N,\mathbf{D}^C)}] \\ &\approx \frac{1}{B}\sum_{1}^{B} h(\mathbf{N}^b,\mathbf{C}^b,\theta^b)w(\mathbf{N}^b,\mathbf{C}^b,\theta^b). \end{aligned}$$

where $w(\mathbf{N}^b,\mathbf{C}^b,\theta^b) = \frac{\pi(\mathbf{N}^b,\mathbf{C}^b,\theta^b|\mathbf{D}^N,\mathbf{D}^C)}{q(\mathbf{N}^b,\mathbf{C}^b,\theta^b|\mathbf{D}^N,\mathbf{D}^C)}$ are the importance weights. Note that is we use Method 1, θ^b is the same for all b. The importance weights vary depending on which proposal distribution we choose. However, there exists a proposal distribution which results in importance weights equal to $\pi(\mathbf{D}^N|\mathbf{N},\theta)$, which we have assumed to be a binomial distribution. Luckily, we know this proposal distribution and can sample from it. It was Geir Storvik who first stated and proved the following theorem, in an unpublished note.

Theorem: (Geir Storvik)

Theorem 5.3.1. *If we choose the proposal distribution*

$$q(\boldsymbol{N},\boldsymbol{C},\theta|\boldsymbol{D}^N,\boldsymbol{D}^C) = \pi(\boldsymbol{N}|\boldsymbol{C},\theta)\pi(\boldsymbol{C}|\boldsymbol{D}^C,\theta)\pi(\theta),$$

then the importance weights reduce to

$$w(\boldsymbol{N}^b,\boldsymbol{C}^b,\theta^b) \propto \pi(\boldsymbol{D}^N|\boldsymbol{N}^b,\theta^b).$$

Proof. First we use Bayes rule to see that

$$\begin{aligned} \pi(\mathbf{N},\mathbf{C},\theta|\mathbf{D}^N,\mathbf{D}^C) &\propto \pi(\mathbf{D}^N,\mathbf{D}^C|\mathbf{N},\mathbf{C},\theta)\pi(\mathbf{N},\mathbf{C},\theta) \\ &\propto \pi(\mathbf{D}^N|\mathbf{N},\theta)\pi(\mathbf{D}^C|\mathbf{C},\theta)\pi(\mathbf{C}|\theta)\pi(\mathbf{N}|\mathbf{C},\theta)\pi(\theta) \\ &\propto \pi(\mathbf{D}^N|\mathbf{N},\theta)\pi(\mathbf{D}^C|\mathbf{C},\theta)\pi(C|\theta)\pi(\mathbf{N}|\mathbf{C},\theta)\pi(\theta) \\ &\propto \pi(\mathbf{D}^N|\mathbf{N},\theta)\pi(\mathbf{N}|\mathbf{C},\theta)\pi(\mathbf{C}|\mathbf{D}^C,\theta)\pi(\theta). \end{aligned}$$

Then, with the proposal function $q(\mathbf{N}, \mathbf{C}, \theta|\mathbf{D}^N, \mathbf{D}^C) = \pi(\mathbf{N}|\mathbf{C}, \theta)\pi(\mathbf{C}|\mathbf{D}^C, \theta)\pi(\theta)$ the weights now reduce to

$$\begin{aligned} w(\mathbf{N}^b, \mathbf{C}^b, \theta^b) &= \frac{\pi(\mathbf{N}^b, \mathbf{C}^b, \theta^b|\mathbf{D}^N, \mathbf{D}^C)}{q(\mathbf{N}^b, \mathbf{C}^b, \theta^b|\mathbf{D}^N, \mathbf{D}^C)} \\ &\propto \frac{\pi(\mathbf{D}^N|\mathbf{N}, \theta)\pi(\mathbf{N}|\mathbf{C}, \theta)\pi(\mathbf{C}|\mathbf{D}^C, \theta)\pi(\theta)}{\pi(\mathbf{N}|\mathbf{C}, \theta)\pi(\mathbf{C}|\mathbf{D}^C, \theta)\pi(\theta)} \\ &\propto \pi(\mathbf{D}^N|\mathbf{N}, \theta). \end{aligned}$$

□

This proposal distribution is known to us. The (prior) distribution of the model parameters are specified as in Method 1 or Method 2. Further, we spent almost the entire Chapter 3 to find the distribution of $\mathbf{N}|\mathbf{C}, \theta$. Finally, we have available posterior samples of $\mathbf{C}|\mathbf{D}^C$ from the *ECA*-program by Hirst et al. (2004). It is a problem that we only have these samples, and not samples of $\mathbf{C}|\mathbf{D}^C, \theta$. We have not come up with any better solution than to simply use $\mathbf{C}|\mathbf{D}^C$ as an approximation of $\mathbf{C}|\mathbf{D}^C, \theta$. If the data $\mathbf{D}^C$ are very informative, the approximation should be good.

5.3.2 Sequential importance sampling with resampling (SISR)

In the SISR-algorithm we draw our samples *sequentially*, i.e. year by year. These samples are often called *particles*. After we have obtained particles for a whole year, we can pause and might choose to *resample*. We resample the particles we have obtained with replacement, with probabilities equal to the normalized importance weights w_y^b. Then particles with large weights are duplicated, and those with very small weights will probably disappear. When the resampling is done, the information in the weights are sort of "used", and we set all the weights equal to $\frac{1}{B}$. The SISR algorithm is described in Algorithm 5.3.1.

Note that we do not resample at every step. Whether we choose to resample or not depends on the effective sample size. If the ESS is very small, then many weights are close to zero. When these weights are multiplied by weights in other years, they will be even closer to zero. The resampling will rid us of the particles with these small weights. A common choice is to resample when the effective sample size is about half the number of particles (Creal, 2012). If we never resample, the SISR-algorithm reduces to the normal importance sampling.

Algorithm 5.3.1 SISR

1: **for** b=1,...,B **do**

2: Draw $(\mathbf{N}_Y^b, \mathbf{C}_Y^b, \theta_Y) \sim q_Y(N_Y, C_Y, \theta_Y | \mathbf{D}^N, \mathbf{D}^C)$

3: Set $w_Y^b = \pi(D^N | \mathbf{N}_Y^b, \theta_Y^b)$

4: **for** $y = Y-1, Y-2, ..., 1$ **do**

5: **for** $b = 1, ..., B$ **do**

6: Draw $(N_y^b, C_y^b, \theta_y) \sim q_y(N_y, C_y, \theta_y | N_{y+1}^b, C_{y+1}^b, D^N, D^C)$ and compute the importance weights $w_y^b \propto w_{y+1} \pi(D^N | \mathbf{N}_y^b, \theta_y^b)$

7: **for** $b = 1, ..., B$ **do**

8: Normalize the importance weights $\hat{w}_y^b = \frac{w_y^b}{\sum w_y^b}$

9: Calculate the effective sample size $ESS_y = \frac{1}{\sum (w_y^b)^2}$

10: **if** $ESS_y < 0.5 \cdot B$ **then**

11: Resample B particles $(\tilde{N}_y^b, \tilde{C}_y^b, \tilde{\theta}_y^b)$ with probabilities $\{\hat{w}_y^b\}$ and set $w_y^b = \frac{1}{B}$ for all b.

12: **return** $(\tilde{N}, \tilde{C})$

Weights on a log-scale

Since many of these weights may be very large or very small, we will compute them on a log scale to avoid numerical difficulties. Then we obtain $\log w^b, b = 1, ..., B$. We can then normalize these weights by

$$w(N^b, C^b) = \frac{e^{\log w^b}}{\sum_1^B e^{\log w^b}}$$

However this is also numerically unstable since the denominator can be very large. To fix this we just multiply by 1 in the following convenient way.

$$w(N^b, C^b) = \frac{e^{\log w^b - \max \log w^b}}{\sum_1^B e^{\log w^b - \max \log w^b}}.$$

It was Geir Storvik who suggested to calculate the weights in this way.

5.3.3 Finding posterior quantiles

In the literature we were not able to find a formal procedure to find posterior quantiles in the importance sampling setting. A simple solution is to sample from the particles with replacement, with probability equal to the normalized importance weights. Then we can

find quantiles of these data, in the usual way. However, we were able to develop a more direct and precise estimation of the posterior quantiles.
Recall that the q'th quantile n_q is the number such that

$$P(N \leq n_q) = q.$$

However, this probability can be written as an expectation using the indicator function, i.e. $E[\mathbb{I}_{\{N<n_q\}}] = P(N \leq n_q)$. Thus we can use the importance sampling to estimate the quantiles by ordering the abundances from smallest to largest, and letting $h(N, C) = \mathbb{I}_{\{N<n_q\}}$.

$$\begin{aligned} E[\mathbb{I}_{\{N<n_q\}}] &\approx \sum_{N^b} \mathbb{I}_{\{N^b<n_q\}} w(N^b, C^b) \\ &= \sum_{\{N^b<n_q\}} w(N^b, C^b). \end{aligned}$$

Now we set this sum equal to some quantile q, and solve for n_q. In Algorithm 5.3.2 we show how this can be done, and the corresponding R-function is given in Appendix D. Of course, the exact same procedure can be used to find the quantiles of $\mathbf{C}|\mathbf{D}^N, \mathbf{D}^C$, or any of the model parameters.

Algorithm 5.3.2 Finding quantiles

```
1: function FindQuantile(Nb,w,q)
2:     df = data.frame("Nb"=Nb,"w"=w)
3:     ordered = order df ascending by Nb
4:     weight.sum = 0
5:     i = 0
6:     while weight.sum ≤ q do
7:         i = i+1
8:         weight.sum = weight.sum + w[i]
       n_q = Nb[i]
9:     return n_q
```

5.4 The R-functions `StockSizeIS` and `StockSizeISprior`

Based on the methods described in this chapter, we created two R-functions, which we will describe now.

`StockSizeIS`

This function combines the parameter estimation of Method 1 with sequential importance sampling. That is, a combination of Algorithms 5.1.1 and 5.3.1. Indices, catches and a catchability array must be given as input to the function. The outputs are estimates of the model parameters $\theta = \{m_{y,a}, f_{y,a}, \lambda_{y,a}\}$, the importance weights $\{w^b\}$, the effective sample size and posterior samples of the abundances. The samples are drawn sequentially.

`StockSizeISprior`

This function combines the prior distributions of Method 2 with SISR, that is Algorithms 5.2.1 and 5.3.1. This function only needs the indices and catches as input. In addition to $\{q_a\}$, $\{w^b\}$ and ESS, the function returns posterior samples of $\mathbf{N}$, $\mathbf{C}$, $\mathbf{m}$, $\mathbf{f}$ and λ.

In both functions we have added an ad hoc restriction on the λ's. If the estimated λ happens to be smaller than the catch, we run into numerical and intuitional difficulties. If this happens, we adjust the λ so that it becomes larger than the catch. The R-code for these functions can be found in Appendices F and G.

Chapter 6

Simulation experiment

Statistics is the grammar of science.

Karl Pearson

To understand the algorithms developed in the previous chapter better, it is very useful to test them on a simulated data set. When we simulate a data set, we know the values of all the model parameters. Thus, we can check if our methods are able to make correct inference. This chapter contains four sections; one about simulating the data set, one about testing of Method 2 and importance sampling, one about testing of Method 2 and importance sampling, and finally a summary section.

6.1 Simulated data set

The relevant R-code can be found in Appendix E.

The data we create should be similar to the real data set described in Chapter 2. Hence, we need some indices and catches. In addition to this, we will simulate the abundances $\mathbf{N}$ and all the model parameters $\theta = \{\lambda, \mathbf{m}, \mathbf{f}, \mathbf{q}\}$. As we recall from Chapter 5, Method 2 involves specifying a lot of prior distributions. Hence we define the simulated data in such a way that these prior distributions do not miss completely. Also, we will make the data set simple, in the sense that we avoid making the differences over the years and ages too complex. Therefore both the natural mortality and the catchability will be set constant over years and age groups. However, we let the fishery mortality be different between the age groups, to match the prior distributions we use in Method 2.

When values are assigned to all the model parameters, we create $B = 5\ 000$ samples of $\{N_{y,a}\}$, $\{C_{y,a}\}$ and $\{I_{y,a}\}$. We use the same distributional assumptions on these variables as

those described in the model assumptions in Chapter 3. With these fake data we simulate the parameter estimations of Method 1 or Method 2, together with the estimates of the abundance through importance sampling. This simulation experiment is then repeated $M = 100$ times.

Algorithm 6.1.1 describes how we execute the simulation experiment.

Algorithm 6.1.1 Simulation experiment

1: Let the catchability be constant $q_a = 1.5 \cdot 10^{-7}$.

2: Let the fishery mortality equal the ICES estimates,

$$f = (0.005, 0.034, 0.095, 0.161, 0.256, 0.306, 0.338, 0.304).$$

3: Let the natural mortality equal $m_a = 0.18$ for all a.

4: Let $\lambda_{a=3} = 5 \cdot 10^9$.

5: Calculate $\lambda_{a+1} = \lambda_a(1 - m_a)(1 - f_a)$.

6: Use these values of $\mathbf{q}_a$, $\mathbf{f}_a$, $\mathbf{m}_a$ and λ_a in all years.

7: **for** m in 1 to M **do**

8: Draw $N^m_{y,a}$ from Poisson$(\lambda_{y,a})$.

9: **for** n in 1 to B **do**

10: Draw $C^{b,m}_{y,a}$ from Binom$(N^m_{y,a}, f_{y,a}(1 - m_{y,a}))$.

11: Draw $I^{b,m}_{y,a}$ from Binom$(N^m_{y,a}, q_a)$.

12: Apply either `StockSizeIS` or `StockSizeISprior`, with $C^m_{y,a}$ and $I^m_{y,a}$ (and q_a) as input.

13: Save the estimated values of $m^m_{y,a}$, $f^m_{y,a}$ and $N^{b,m}_{y,a}$, along with the effective sample size.

Remark In Algorithm 6.1.1 we claim to draw samples from the Poisson and binomial distributions. However, this is only approximately true. We observe that the R-functions `rpois` and `rbinom` do not allow for λ or N, respectively, to be much larger than about $2 \cdot 10^9$. If this happens, we simply use the normal approximations

$$\text{Poisson}(\lambda) \approx \text{Normal}(\lambda, \sqrt{\lambda}),$$
$$\text{Binomial}(n, p) \approx \text{Normal}(np, \sqrt{np(1-p)}.$$

6.2 Testing Method 1 and importance sampling

The R-function `StockSizeIS` combines the parameter estimation of Method 1, with the sequential importance sampling with resampling (SISR) procedure described in Algorithm 5.3.1. However, we do not need the resampling part here. With the simulated data sets at hand, we both wish to see how the parameter estimation works, and how the importance sampling works. Since we know the true value of the catchability, we give this as input to our function. The entire R-code for `StockSizeIS` can be found in Appendix F.

We replace line 12 in Algorithm 6.1.1 with

R script

```
test1 <- StockSizeIS(Ind_sim[,,,m],C_sim[,,,,m],q_sim)
```

The effective sample size

For each of the $M = 100$ simulations, we store the effective sample size. As we can see in the histogram of Figure 6.1, the effective sample size is consistently very high, at about 4995.

According to the high effective sample size, no values are "unreasonable". But are the parameter estimates accurate?

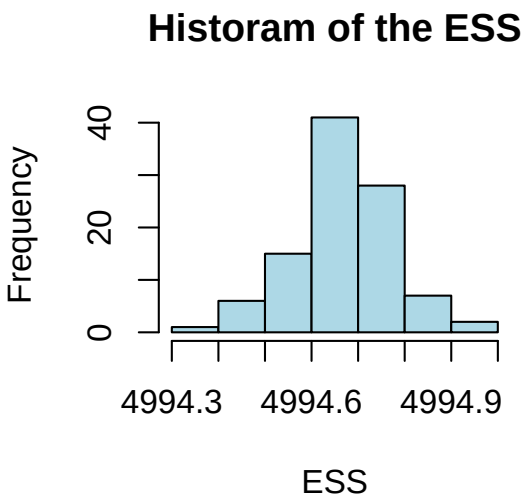


Figure 6.1: The ESS is always about 4995.

Parameter estimation

In this experiment we have assumed the catchability to be known, and have not made any inference about it. However, we want to see if the natural and fishery mortalities are correctly estimated. To determine this, we plot the true values together with the estimated values. Of course, the estimates are a bit different in the $M = 100$ experiments,

but we will not draw 101 lines in each plot. The plots in Figure 6.2 contain only four lines; the true value and the mean, minimum and maximum of the estimates values. In the plots at the left, the mortalities are plotted against the years, and in the plots to the right, they are plotted against the ages.

The natural mortalities are estimated very accurately, as the values only ranges from about 0.178 to 0.182. It is interesting to note that the estimates are less certain for older age groups than for younger. To understand this, we recall how the natural mortalities are estimated,

$$\widehat{m}_{y,a} = \frac{\frac{I_{y,a}}{q_a} - C_{y,a} - \frac{I_{y+1,a+1}}{q_{a+1}}}{\frac{I_{y,a}}{q_a}}.$$

For large age groups, the indices tend to be a lot higher than for younger age groups. In a fraction where the indices are large, the value of the catch does not influence the ratio that much. Hence, the variability in the catches are less visible in estimates for the larger age groups.

The plots of the fishery mortalities looks a bit strange. In the bottom left plot, we see that there is little variation in the estimates here as well. However, we overestimate the fishery mortalities (though, only with about 0.007)! The reason becomes apparent in the bottom right plot. The fishery mortality is correctly estimated for every age group, except the $A+$ group. We miss here because we have not actually estimated $f_{y,A}$. Recall that the estimator of the fishery mortality requires the index in the next year of the next age group, in the following way:

$$\widehat{f}_{y,a} = \frac{C_{y,a}}{\frac{I_{y+1,a+1}}{q_{a+1}} + C_{y,a}}.$$

Because age group $A+1$ does not exist, we simply assumed that the fishery mortality in the $A+$ group equals the mortality in age group $A-1$. Here we suffer the consequence. On the other hand, the suffering is tolerable, and we just have to keep in mind that the mortality for ages $A+$ may be different from $A-1$, even though we can not see it in our results.

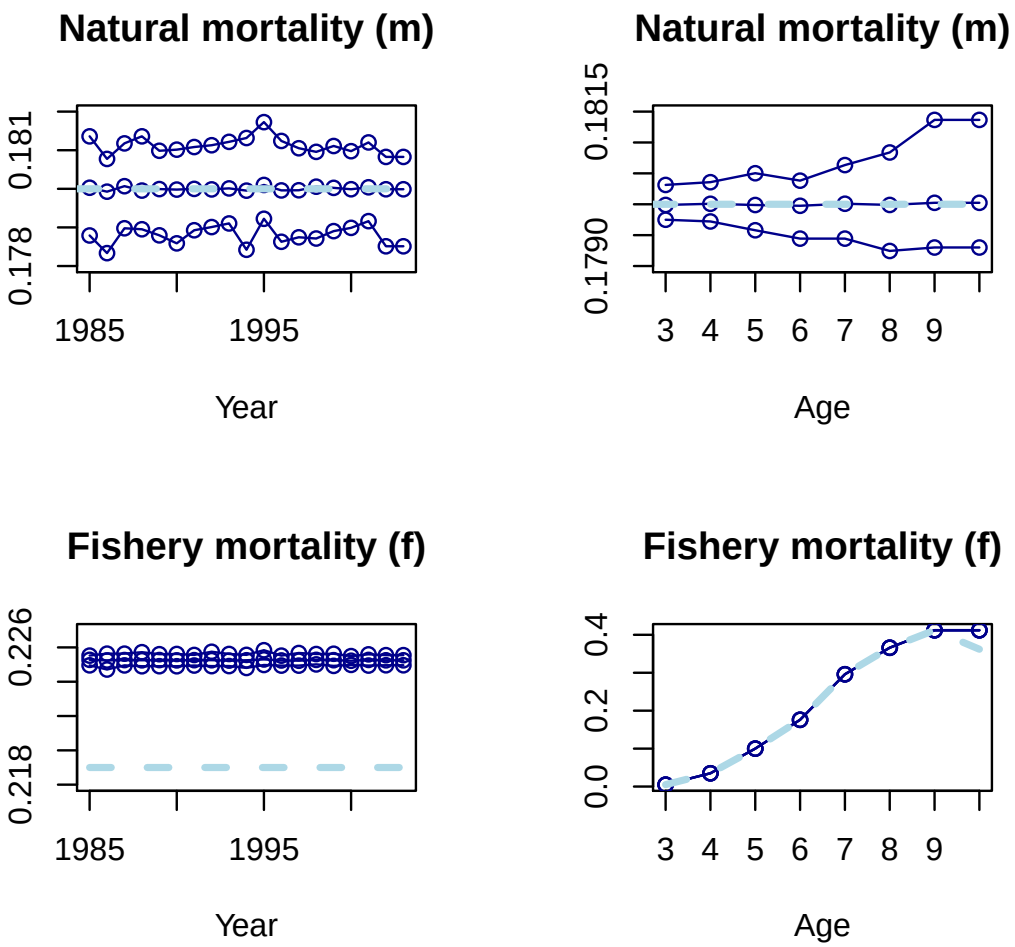


Figure 6.2: The light blue colour represents the true value. In the bottom left plot, the "true value" is the mean over all age groups. The dark blue lines are the mean, minimum and maximum of the estimated values among the 100 experiments.

Posterior abundance

Now we are reassured that the parameter estimation works well, and we are eager to see if the importance sampling procedure gives reasonable inference about the abundances. To test this we calculate the normalized difference between the true simulated values and the mean of the particles. That is

$$D_{y,a}^{m} = \frac{\frac{1}{B}\sum_{b=1}^{B}\left[N_{y,a}^{\text{estimate,b,m}}\right] - N_{y,a}^{\text{simulated, m}}}{N_{y,a}^{\text{simulated,m}}}.$$

Figure 6.3 displays the average difference plotted against the years and ages. From this we see that our estimates are actually too low. It is not clear why this happens, but one possible explanation lies in the fishery mortalities. When the fishery mortalities are too high, this will lead us to lower estimates of the abundances. However, considering the magnitude in the overestimation of the fishery mortality, there must be something else going on here.

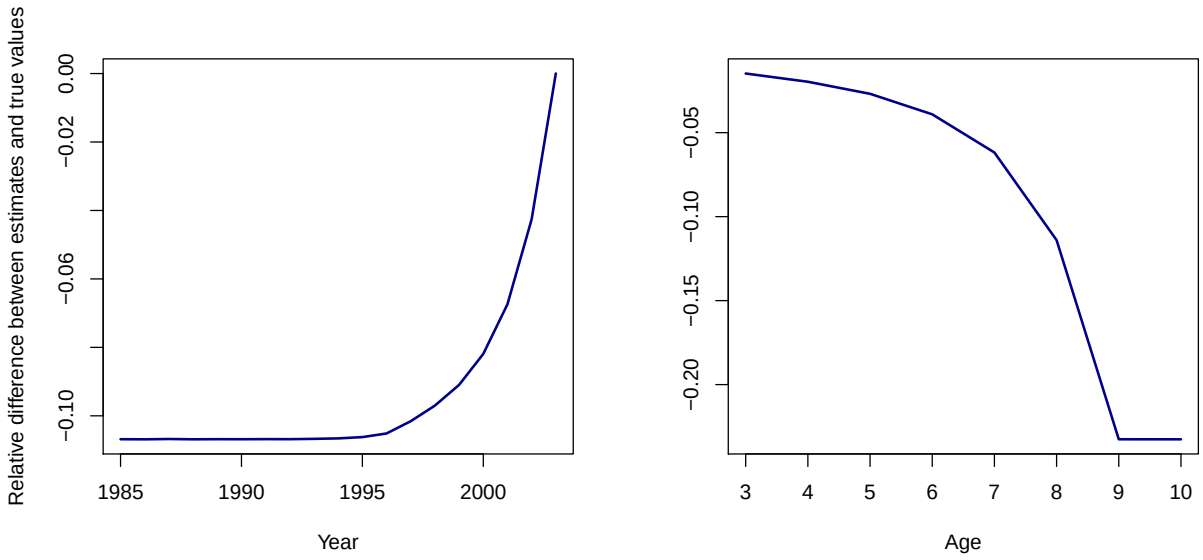


Figure 6.3: Relative difference between estimated values and true values.

6.3 Testing Method 2 and importance sampling

We proceed to investigate how the Bayesian program works. `StockSizeISprior` combines the parameter estimation of Method 2 with SISR. Further, `StockSizeISprior` runs sequentially like `StockSizeIS`, but now we will also need the *resampling* idea of SISR. The entire R-code for this function can be found in Appendix G. Again we run the function using the simulated data set described in Section 6.1. We replace line 12 in Algorithm 6.1.1 with

R script

```
test2 <- StockSizeISprior(Ind_sim[,,,m],C_sim[,,,,m])
```

In this case, we store a $Y \cdot S \cdot A \cdot N$ dimensional array of fishery and natural mortalities in each of the $M = 100$ simulations. Already when we reach $m = 2$, a complaint about the memory usage appears. To be able to perform this experiment, we let $B = 500$ instead of $B = 5000$. Because of this, we expect less accurate results than in the previous section.

The effective sample size

The effective sample size is always close to 1. The histogram in Figure 6.4 illustrates this.

We interpret a low effective sample size to mean that the posterior samples do not match the data. This is very strange because the prior distributions we use are all centred around the true simulated values. The problem may be that we have *too many* parameters. Our algorithm aims to sample from the distributions of $Y \cdot A$ natural mortalities, $Y \cdot A$ fishery mortalities, $Y \cdot A$ intensities, $Y \cdot A$ abundances and $Y \cdot A$ catches! That is a total of

$$5 \cdot Y \cdot A = 5 \cdot 19 \cdot 8 = 760$$

parameters. When we have 760 possibilities to draw an unlikely value, then it is not so strange that we do. With just one effective sample, we do not even bother to show the results of the analysis. However, we can mention that all the posterior distributions looked just like the prior distributions.

After many attempts to fix the algorithm, we were forced to give up.

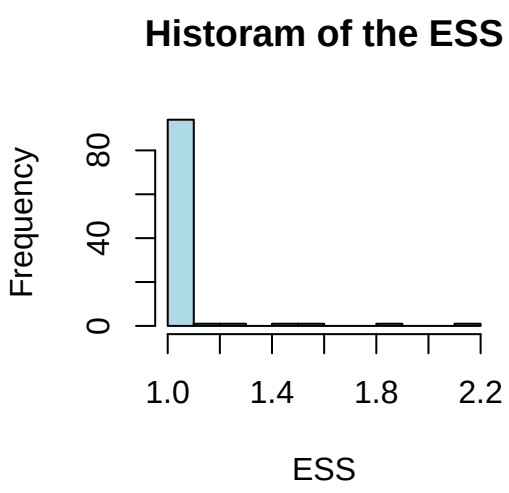


Figure 6.4: Effective sample size in the 100 experiments when $B = 500$.

6.4 Summary of the testing

The first algorithm that uses GMM estimation runs smoothly and gives reasonable output, but there were some problems with underestimation of the abundance. However, this was not of a large magnitude, and may partly be explained by a reasonable overestimation of the fishery mortalities.

When it comes to Method 2, the test results are not satisfying. One idea is to run the function with more particles. Here, we have only used $B = 500$, because this is what the author's laptop memory can handle. However, the running time is only a few seconds, and thus it would be possible to use, say, 50 million particles and some hours on a supercomputer. That should give us a higher effective sample size, and maybe more interesting output. The task of using a supercomputer is left on the to-do-list for future research.

For this thesis project, we give up on Method 2 and `StockSizeISprior`. When the algorithm does not work in a perfect simulated world, there is no point in applying it to the real data set. Hence, when we in the next chapter present the analysis of the real data, we will only use Method 1, and the R-function `StockSizeIS`.

Chapter 7

Results

Latin: To spring back, rebound, equivalent to *re-* + *-sultāre*, combining form of *saltāre* to dance (frequentative of *salīre* to leap, spring).

Origin of the word *result*(Result, 2015)

In this chapter we present the results of analysing the data described in Chapter 2. We use the function `StockSizeIS` to analyse the data, which combines the parameter estimation of Method 1 with importance sampling. This function was tested on a simulated data set in the previous chapter, and we are quite confident that it will give us reasonable parameter estimates. However, in the simulated data set we knew the value of the catchability $\{q_a\}$. This is not the case in the real world, so we have to try out different possibilities.

In the first section we let q_a be constant over the age groups. This makes the results easy to interpret, because the catchability then is a pure scaling of the total abundance. However, we do not believe that the catchability really is constant. In Section 7.2, we try to use the catchabilities we calculated in Chapter 2. These were estimated to make the indices match the ICES-estimates. Because we have not found any way to estimate the catchability, this seems to be our best guess at what the catchabilities might be. Readers with better ideas of what the catchability might be, are free to try out the R-function and see what happens. To run the function we write:

R script

```
run <- StockSizeIS(Ind,C,q)
```

7.1 Constant catchability

Here we assume that $q_a = q$ for all ages a. We run `StockSizeIS` with the catch and index data described in Chapter 2. We have decided to run the importance sampling procedure with $B = 5\ 000$ particles. This choice is based on two wishes:

1. We want many samples to make good inference, thus we desire a high value of B.
2. We do not want the computer algorithm to use too much time, nor crash because of memory shortage, and hence we must set a limit for B somewhere.

In order to run the particle filter with $B = 5\ 000$ particles, we need to input 5 000 catches per age-year group. This is because there must be one set of catches corresponding to each particle of abundances. In the data set we have available, which is described in Chapter 2, 500 samples of the catch in every age-year group are given. Now we repeat every catch observation ten times, which gives us $B = 5\ 000$ samples from approximately $\pi(\mathbf{C}|\mathbf{D}^C)$.

When q is constant over ages it is easy to see how the catchability scales the system. To make comparison with the next section easier, we let the constant catchability equal the mean of the catchabilities that match the ICES estimates. That is

$$\begin{aligned} q &= \frac{1}{A}\sum_{a=1}^{A} q_a^{(ICES)} \\ &\approx 5.46 \cdot 10^{-7}. \end{aligned}$$

When we run the R-function with this catchability, the effective sample size is about 109, which is only about 2% of the number of particles.

R script

```
> run$ess
[1] 109.27
```

However, with more than 100 effective samples, we should be able to make some inference. We consider the parameter estimates first, and then the abundances.

Parameter estimates

In this section we look at the estimates of the mortalities $\{m_{y,a}\}$ and $\{f_{y,a}\}$. In Figure 7.1 we see the mean estimates for different years and for different ages. The bottom right plot

shows that there is a higher probability of being fished for older, and thus larger, NEA cod. This seems reasonable since they are too large to escape through the nets of the trawl.

Turning to look at the top right figure we see that the natural mortality is lowest for the eldest, but is also low for the youngest NEA cod. This is not what we would expect. The youngest fish are more exposed to predation, and should thus have the highest mortality. It would also be reasonable if the eldest fish had a higher mortality. After all, old age is associated with a larger probability of dying in most species.

In the top left plot we note that the natural mortality was high in 1989 and 1998. There is quite large variation between the years, as the natural mortality ranges from about 10% to 30%. Some of this variation can be explained by the fact that the mean age of the population varies. Hence, we are led to believe that many recruits joined the population the years before 1989 and 1998. The fishery mortality, displayed in the bottom left plot, follows a curve similar to the catches, displayed in Chapter 2. We recognise the restrictions set on the fisheries in the early '90s.

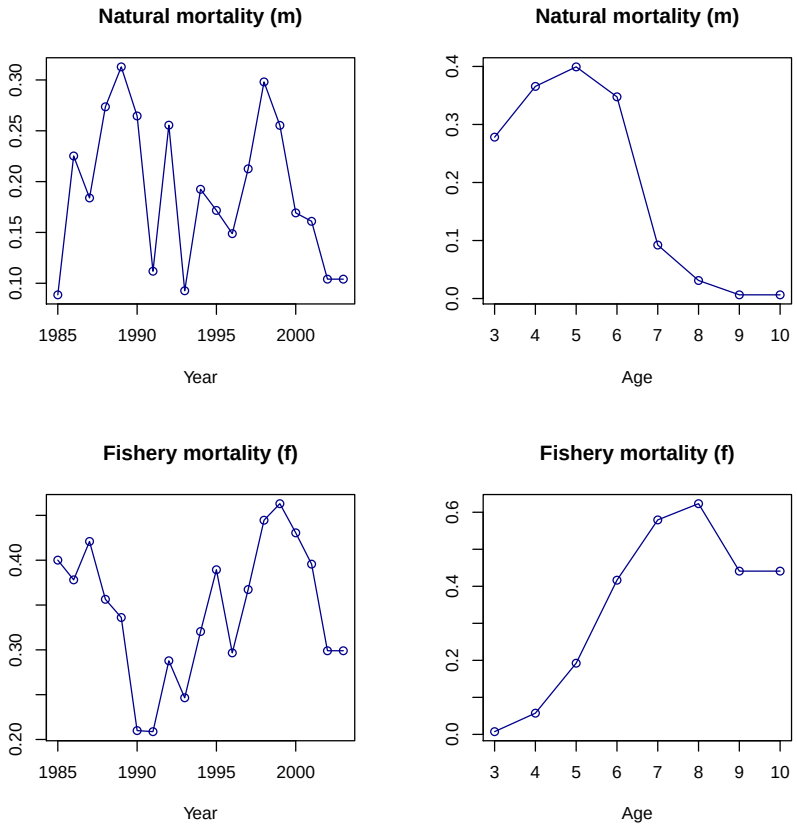


Figure 7.1: The mean mortalities over ages and years. Case when $q = 5.46 \cdot 10^{-7}$.

Posterior distribution of abundance and catch

The posterior mean, median, 5% and 95% quantiles of both the catches and the abundances are shown in Figure 7.3.

Let us first have a look at the abundances. Even though we have about 100 effective samples, the 5% and 95% quantiles seems to be glued together. Actually, the same is true if we look at the maximum and minimum! This is because there is little variation in the Poisson distribution, which we draw our samples from. The variation is at least small compared to the changes from year to year. Also, since the catch is less than one tenth of the abundance, the influence of the variation in the catch is not visible in the plot of the abundance. However, this does not mean that we are certain of these numbers. These posterior distributions are based on our estimated and then fixed parameter values. The uncertainty really lies in the parameters, not in the model.

Posterior abundance vs. the indices

The posterior abundance looks very much like the plot of the indices in Chapter 1. Since we only use the indices to estimate λ, we want to check if the catches actually influenced the results at all. Figure 7.2 shows the posterior mean along with the estimated λ's. These λ's are just a scaling of the indices. We see that the posterior mean is very similar to the λ's, which suggests that our results are overfitted to the index data.

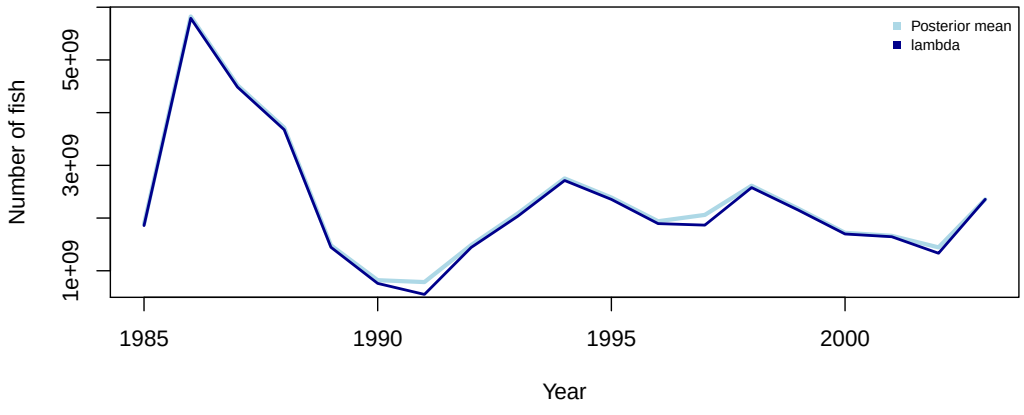


Figure 7.2: The posterior mean along with the estimated λ's

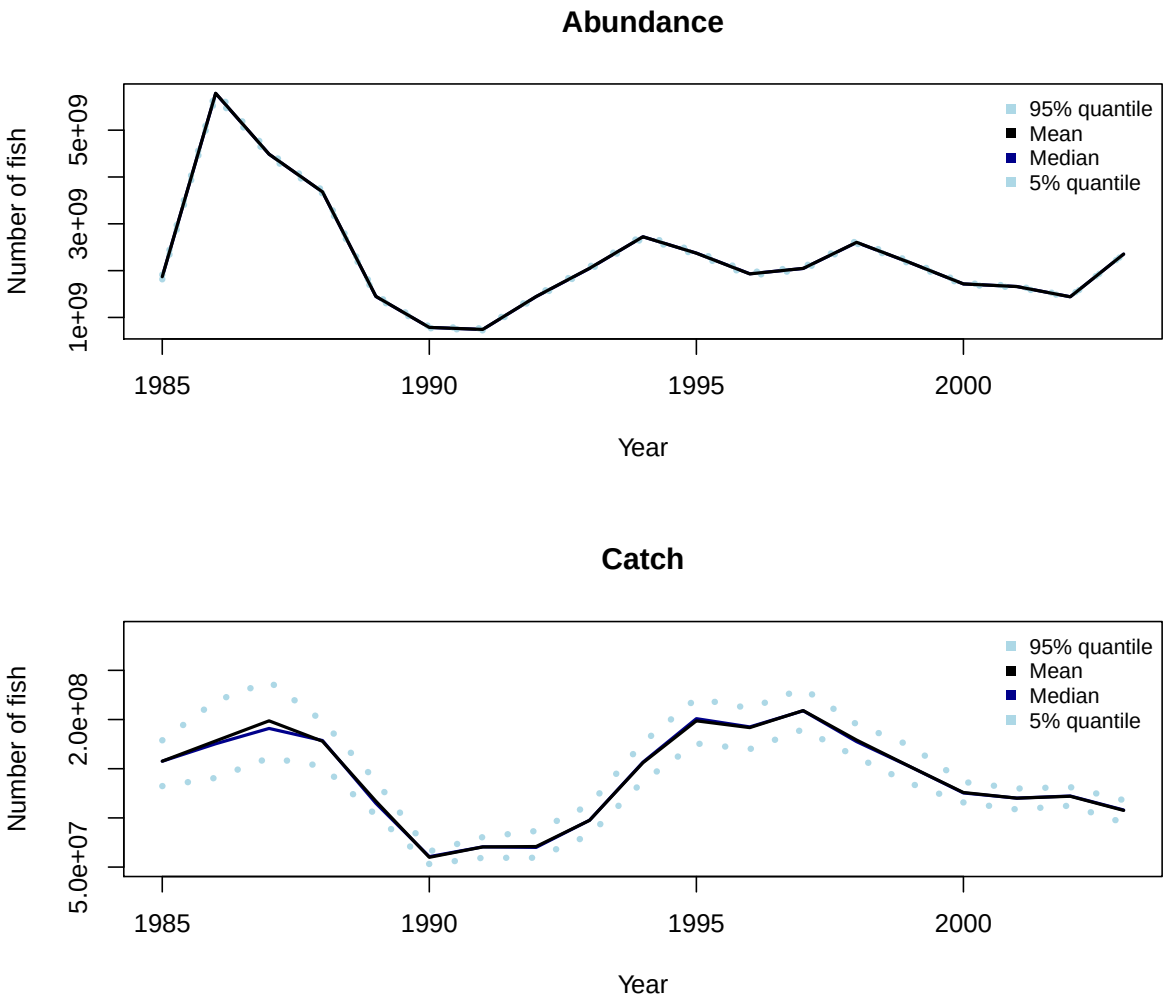


Figure 7.3: Here are the posterior median and 5% and 95% quantiles of both the abundance and catches. Here we have used $q = 5.46 \cdot 10^{-7}$.

When it comes to the catch, the uncertainty is visible. The mean is very similar to the mean we plotted in Chapter 2, which indicates that the mean of $\mathbf{C}|\mathbf{D}^C$ is not very affected when $\mathbf{D}^N$ is also known. When we compare

$$\pi(\mathbf{C}|\mathbf{D}^C)$$

with

$$\pi(\mathbf{C}|\mathbf{D}^C, \mathbf{D}^N),$$

we also wish to know if the uncertainty in the two distributions are similar. Figure 7.4 shows the median, 5% and 95% quantiles of both of these distributions, in the indicated colours. Over all, the distributions look quite similar, but we note two differences. Firstly, the estimates are slightly lower when *both* data sources are given. Secondly, there is less variation in the estimates when both data are given. That is; with more data, we become more certain about the estimates.

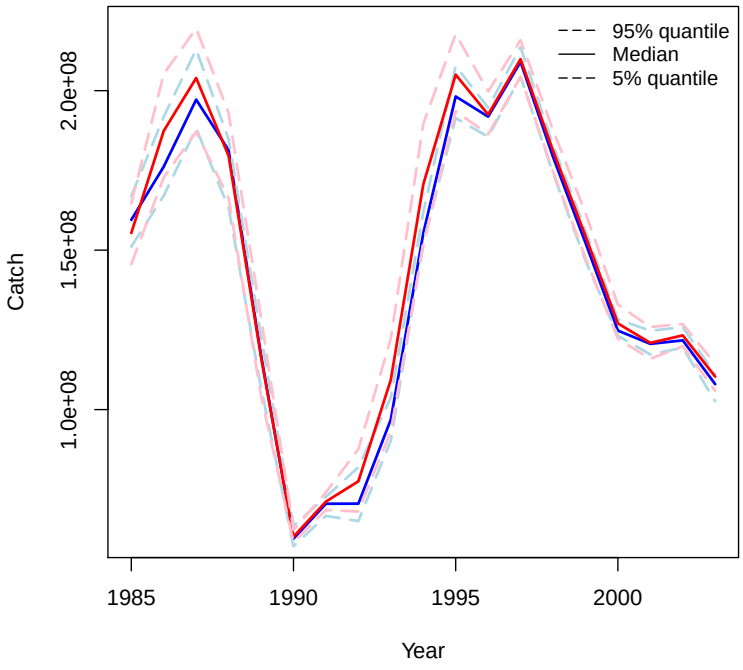


Figure 7.4: Posterior mean, 5% and 95% quantiles of the catch. The red and pink are from $\mathbf{C}|\mathbf{D}^C$, while the blue lines are from $\mathbf{C}|\mathbf{D}^N, \mathbf{D}^C$.

7.2 Catchability to match ICES-estimates

In this section we give a different catchability array as input. This is the catchability that best fits the ICES estimates, and is calculated by the following equation

$$q_a^{(ICES)} = \frac{\frac{1}{Y}\sum_{y=1}^{Y} I_{y,a}}{\frac{1}{Y}\sum_{y=1}^{Y} N_{y,a}^{(ICES)}}$$

$$\approx 10^{-7} \cdot (9.6, 9.2, 7.5, 6.0, 4.4, 3.1, 2.3, 1.6).$$

Here, $N_{y,a}^{(ICES)}$ are the ICES estimates and $I_{y,a}$ are the indices. These catchabilities were plotted in Chapter 2, but we include a small plot here as well, to recall what it looks like. On average, these catchabilities equal the value we used in the previous section, namely $5.46 \cdot 10^7$. Hence, the results here should be on the same level, but the dynamic might look a bit different.

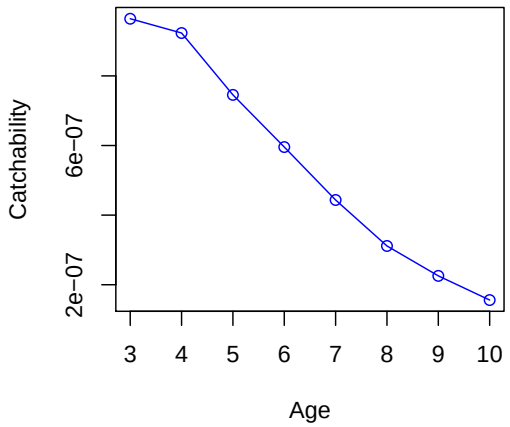


Figure 7.5: Catchabilities.

With this catchability as input the effective sample size is very low, at about 27. This is not much, considering that we started out with 5000 particles.

We start to investigate the estimates of the mortalities, and then look at the posterior distribution of the abundances and the catches.

Parameter estimates

Figure 7.6 shows the plots of the natural and fishery mortalities against ages and years. The bottom plots of the fishery mortalities look quite similar to the plots in the previous section. This is very interesting! Even though we have made a dramatic change in the values of the catchabilities, the fishery mortalities are not very much affected. This is actually a bit strange. Referring to the R-code in Appendix F, and to Theorem 5.1.2, the fishery mortalities are estimated by:

$$\widehat{\overline{N}}_{y,a} = \lambda_{y+1,a+1} + \frac{1}{B}\sum_{b=1}^{B} C_{y,a}^b,$$

$$\widehat{f}_{y,a} = \frac{\frac{1}{B}\sum_{b=1}^{B} C_{y,a}^b}{\widehat{\overline{N}}_{y,a}},$$

where these λ's in turn are estimated using the catchabilities,

$$\widehat{q_a \lambda_{y,a}} = I_{y,a},$$
$$\widehat{\lambda}_{y,a} = \frac{\widehat{q_a \lambda_{y,a}}}{q_a}.$$

It is not clear why the fishery mortality is not influenced more by the change in the catchabilities. However, when it comes to the natural mortalities, these are clearly influenced by the changes in the catchablities. We see a change in the shape of the natural mortalities plotted against age. This plot is more reasonable than the one in the previous section, because the youngest fish have a higher mortality. However, it is still unreasonable that the eldest fish have a mortality close to 0. The estimates in the oldest age groups are sensitive because we often have little data on them, and this might be why we get these strange results.

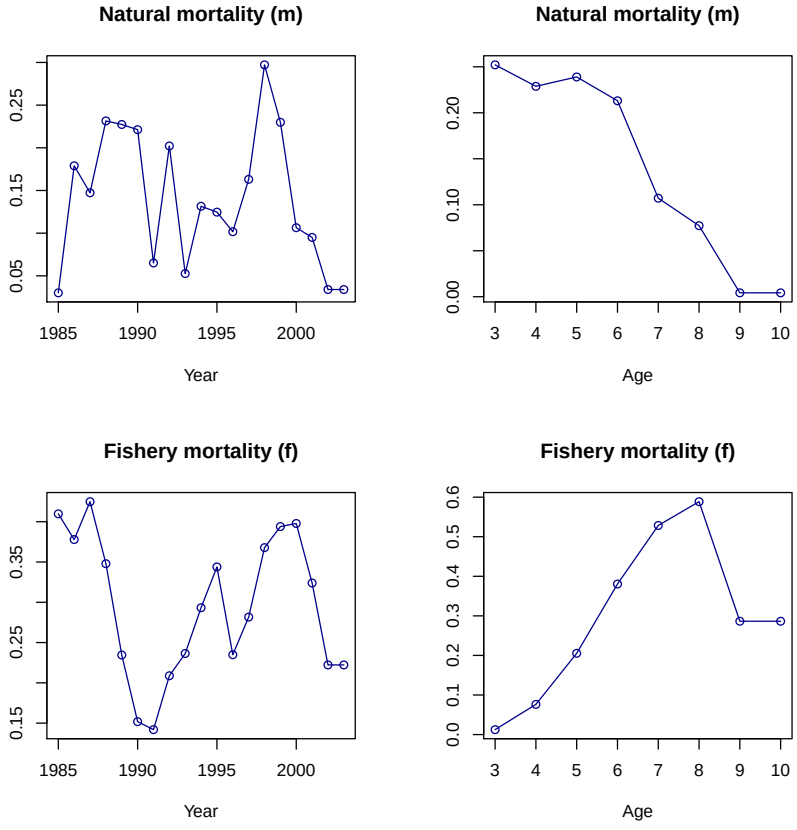


Figure 7.6: The mean mortalities over ages and years. Case when $q = 10^{-7} \cdot [9.6, 9.2, 7.5, 6.0, 4.4, 3.1, 2.3, 1.6]$.

Posterior of abundance and catch

The plots in Figure 7.7 looks surprisingly similar to those in Figure 7.3. Then again, the plots are made by summing over all age groups, and this might be why the total is the same. If this is the case, it suggests that the catchability do not influence the dynamics from year to year, except from the level. However, the abundance in the different age groups are influenced by the catchabilities, and this may in turn be important if our purpose is to predict future years.

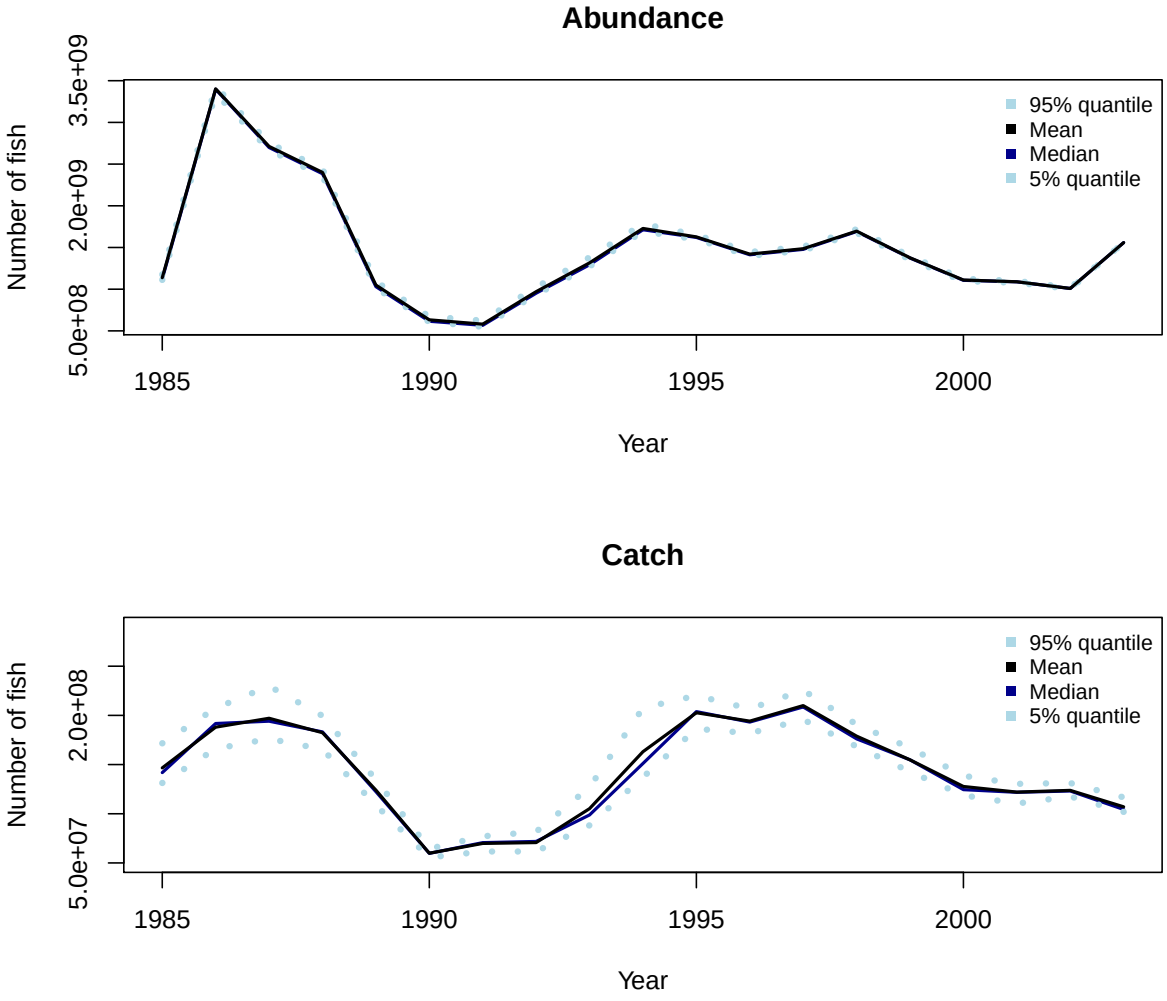


Figure 7.7: Here are the posterior median and 5% and 95% quantiles of both the abundance and catches.

7.3 Summary of the results

The dynamics of the abundances from year to year seems to be a direct image of the indices and our choice of the catchabilities $\{q_a\}$. One might argue then, that it is better to simply estimate the abundance by dividing the indices by some $\{q_a\}$, and not "hide" it inside these complicated algorithms. However, this algorithm gives us a lot of interesting output, which are by-products of the complicated way we estimate the abundances.

Firstly, we obtain estimates of the mortalities $\{m_{y,a}, f_{y,a}\}$. The simulation experiment in Chapter 6 revealed that if the catchability is correctly specified, then these estimates are very precise, and knowledge about the mortalities are very interesting in itself.

Secondly, we obtain samples of the catches given both the catch data $\mathbf{D}^C$ and the research data $\mathbf{D}^N$. This distribution seems to have less variation than the distribution of the catch given only $\mathbf{D}^C$.

As a third point, we noted in this chapter that our uncertainty lies in the parameter estimates, and not so much in the model of the abundances. As we will discuss in the following chapter, it is possible to build a Metropolis-Hastings algorithm on the model parameters, and thus account for this uncertainty.

Chapter 8

Concluding remarks

> I keep saying the sexy job in the next ten years will be statisticians.
>
> Hal Varian

This chapter contains a summary of the thesis and discussion, which also includes suggestions for further work.

8.1 Summary

In this master's thesis we have learned a great deal about the Poisson-binomial model. From a few quite simple model assumptions, we were able to build a model that describes the abundance and catch quite well. A lot of effort was put into exploring the properties of the model. Specially, it was crucial that we managed to show that

$$\pi(\mathbf{N}|\mathbf{C})$$

was equal to a product of probability distributions on the form $\pi(N_{y,a}|C_{y,a}, N_{y+1,a+1})$ (and some more complicated pieces when the last age group was involved). This result made it possible to choose a good proposal distribution for the importance sampling algorithm. The proposal distribution was *good* in two senses:

1. It was easy to obtain samples from.

2. It resulted in simple importance weights.

However, it was not enough to specify a model and an inference algorithm for the abundances and catches. We also needed to consider the model parameters; the natural mortalities, fishery mortalities, catchabilities and the Poisson intensities. In our model, these parameters are necessary to make inference about the abundance, and further, they are interesting *per se*. We investigated the identifiability and estimability properties of the model. It turns out, that is we assume either the catchabilities, fishery mortalities or the Poisson intensities to be known, then everything else is identifiable.

We developed two computer programs to make inference about the model parameters, Method 1 and Method 2. The first uses a frequentistic approach and estimates all model parameters from the data, while the second algorithm uses an Empirical Bayes approach, and puts prior distributions on the parameters. An importance sampling algorithm was built on top of these methods, to obtain Monte Carlo samples from

$$\pi(\mathbf{N}, \mathbf{C} | \mathbf{D}^N, \mathbf{D}^C).$$

Simulation experiments revealed that the second algorithm do not work very well. However, the first algorithm stood the simulation test, and we analysed the real data using Method 1. The results of the analysis are (of course) very dependent on the value we give the catchability. Further, there was little variation in the posterior samples of the abundance. This indicates that the uncertainty really lies in the parameter estimates.

8.2 Discussion and suggestions for further work

There is room for many improvements on the work in this thesis. In this section we briefly describe a few ideas about what can be done in the future.

Divide into seasons

In the model we assume that all the catch is made at the end of the period, after all the natural deaths have occurred. This approximation sounds unintuitive for yearly data, but maybe not that bad for shorter time intervals. This motivates us to introduce a new variable, called *season*. Actually, we started to work with a season-divided model at the beginning of the project, but simplified it because of the bothersome notation, the need of extra results and because it is harder to visualize. As an example, when the model splits into more seasons, the Poisson intensities look like this:

$$\lambda_{y,s,a} = \begin{cases} \lambda_{y,s-1,a}(1 - m_{y,s-1,a})(1 - f_{y,s-1,a}) & s > 1, \\ \lambda_{y-1,S,a-1}(1 - m_{y-1,S,a-1})(1 - f_{y-1,S,a-1}) & s = 1, a < A - 1. \end{cases}$$

In general, everything is split into two cases, when $s = 1$ and when $s > 1$. The difference is that we do not change the age when we move between seasons within the year. The work we have done can be interpreted as a seasonal model with total number of seasons equal to 1. To extend to more seasons, we would have to prove all the results for $s > 1$ as well. In most cases, the arguments will look very similar as those for $s = 1$.

A benefit of this approach is that we might be able to utilise the knowledge about which season the different data are collected.

Test method on Sand Lace

When I attended a one-day seminar at the Norwegian Institute of Marine Research in April, someone mentioned that very good information about the abundance of Sand Lace (*Tobis* in Norwegian) exists. That is, the indices can be assumed to be on the same level as the abundances, which puts us in a setting where the catchability can be assumed to equal 1. It would be interesting to test out our methods in this settings.

Note however that it would involve different distributional assumptions on the indices, as a binomial distribution with $q = 1$ makes little sense. Instead of a binomial distribution, we might use a normal distribution in the following way

$$I_{y,a} \sim N_{y,a} e^{\epsilon_{y,a}},$$

where $\epsilon_{y,a} \sim N(0, \sigma_{y,a})$.

Extension of Method 1: Particle Markov chain Monte Carlo

When we use Method 1, we obtain fixed estimates of the model parameters. Then we use importance sampling to obtain samples from

$$\pi(\mathbf{N}, \mathbf{C} | \mathbf{D}^N, \mathbf{D}^C).$$

However, we would like to account for the uncertainty in the parameter estimates. A suggestion to future students is to build a Metropolis-Hastings algorithm on top of the particle filter. This approach, among other *Particle Markov chain Monte Carlo* (PMCMC) methods are described in Andrieu et al. (2010).

Some thoughts about the catches

In our computer algorithms, the indices had the largest influence. That is because the intensities $\{\lambda_{y,a}\}$ are estimated by the indices alone. However, if we have the choice between using the indices and catches, the indices seem more trustworthy. There is a kind of circular argument involved with using the catch data.

1. Make a guess about the abundance.
2. Set the quota as a percentage of this abundance.
3. In the next year catch data are available. However, the catch is approximately equal to the quota, since the fisheries have no problem harvesting as much as they are allowed to.
4. Estimate the abundance in this year, based on the catch.

One could argue thus, that the catch is simply a reflection of the previous belief about the abundance.

Further work on identifiability

There might me more to reveal about the identifiability of the model. Future students are urged to try and *prove* whether all model parameters can be identified or not. Recall that we showed in Chapter 4 that if the catchability, fishery mortality or intensities were known, then everything else is identifiable. However, we did not say that it is impossible to identify, and maybe also estimate, everything without assuming some parameters to be known. If this is *impossible*, then it would be nice to have a proof of if. On the other hand, if it is *possible*, a proof for that would be great!

Dependency structures in the λ's

Recall from Section 3.1 that we were not entirely comfortable with the binomial assumptions. This was because we doubt that the observations are independent, and that the mortality is the same for each individual fish.

It is probably too complex to account for individual differences, but we can build independence structures between the observations.

Instead of letting the intensities be independent, like in Section 3.1, we could make a linear model

$$\lambda^i_{y,a} = \lambda + \beta_1 a + \beta_2 y + \boldsymbol{\beta} X_i + \epsilon_{y,a}.$$

That is, the intensities depend on the year, age and some other explanatory variables X. X can e.g. contain temperature or location. Linear terms in the age and year is probably too weak. Maybe a better idea would be to have one factor for every year and every age group.

Bibliography

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society*.

Billingsley, P. (1995). *Probability and Measure*. A Wiley-Interscience Publication, 3 edition.

Carlin, B. P. and Louis, T. A. (2009). *Bayesian Methods for Data Analysis - 3rd ed.* Taylor and Francis Group, LLC.

Casella, G. and Berger, R. L. (1990). *Statistical Inference*. Wadsworth and Brooks/Cole Advanced Books and Software, second edition.

Casella, G., Berger, R. L., and Santana, D. (2001). Solutions Manual for Statistical Inference, Second Edition. http://exampleproblems.com/Solutions-Casella-Berger.pdf.

Creal, D. (2012). A Survey of Sequantial Monte Carlo Methods for Economics and Finance. *Econometric Reviews*, 31:3.

Darwiche, A. (2009). *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press.

Hamilton, L. and Butler, M. (2001). Outport Adaptations: Social Indicators through Newfoundland's Cod Crisis. *Human Ecology Review*, 8.

Hirst, D., Storvik, G., Aldrin, M., Aanes, S., and Huseby, R. (2004). Estimating catch-at-age by combining data from different sources. *Can. J. Fish. Aquat. Sci.*, 62.

ICES (2014). Report of the Arctic Fisheries Working Group. Technical report, ICES, Lisbon, Portugal. ICES CM 2014/ACOM:05. 656 pp.

Jakobsen, T. and Ozhigin, V. K. (2011). *THE BARENTS SEA; Ecosystem, Resources and Management*. Tapir academic press.

Lassen, H. and Medly, P. (2001). *Virtual Population Analysis - A Practical Manual for Stock Assessment.* FAO Fisheries Technical Paper 400.

MacKean, P. M., Storvik, G., and Aanes, S. (2015). A State-Space Model for Abundance Estimation from Bottom Trawl Data with Applications to Norwegian Winter Survey. Statistical Research Report 1, Dept. of. math, University of Oslo. ISSN 0806-3842.

Moore, D., Craig, B., and McCabe, G. (2014). *Introduction to the Practice of Statistics.* W.H. Freeman and Company, eighth edition.

Paulino, C. and Pereira, C. (1994). On Identifiability of Parametric Statistical Models. *J. Ital. Statist. Soc.*, 1.

R Core Team (2015). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Result (2015). Dictionary.com unabridged. *http://dictionary.reference.com/browse/result.*

Ross, S. M. (2007). *Introduction to Probability Models.* Elsevier Inc., ninth edition.

Santanu, S. R. (2013). *Graph Theory with Algorithms and its Applications.* Springer.

Shepherd, J. G. (1999). Extended Survivors Analysis: An improved method for the analysis of catch-at-age data and abundance indices. *ICES Journal of Marine Science.*

Ziegler, A. (2011). *Generalized Estimating Equations.* Springer.

Appendix A

Data description

A description of the data file `cod1985.2003.3.15` .

Ind: An array of dimension Y x B x A where $Y = 19$ is the number of years, $B = 100$ the number of bootstrap samples of $I_{y,s^*,a}$ and $A = 13$ the number of age categories. These data are called $\mathbf{D}^N$ in the next chapter.

Indsum: An array of dimension Y x A x 2 giving mean and standard deviations of the indices.

caa: Array of dimension Y x A x S x M where $S = 12$ is the number of seasons and $M = 500$ the number of posterior samples of $\frac{C_{y,s,a}}{10^6}$.

year: A vector giving the years with data - this is just the years $(1985, 1986, ..., 2003)$.

age: A vector giving the age groups - an integer vector from 3 to 15. The first and last group also contain the fish aged below 3 years and above 15 years, respectively.

surveyS: An index giving the season corresponding to survey data, which is just the number 2, i.e. February.

Appendix B

FixingData

```
library(StockEstim)

############################
# Load the cod data
############################

data(cod1985.2003.3.15)
cod <- cod1985.2003.3.15

#########################################################################
# Transpose both arrays to have dimension (Y,A,100) and
# (Y,S,A,500), respectively. We sum the catches over the seasons
# to end up with yearly data. We also multiply the catches by
# 10^6, because caa has earlier been scaled down by this factor.
#########################################################################

Ind2 <- aperm(cod$Ind,c(1,3,2))
C <- array(apply(10^6*aperm(cod$caa, c(1,3,2,4)),c(1,3,4),sum),dim=c
    (19,1,13,500))

#####################################################################
# We make a=10 the largest age group, instead of 15.
#####################################################################

Ind10 <-Ind2
Ind10[,8,] <- apply(Ind2[,8:13,],c(1,3),sum)
Ind10 <- Ind10[,1:8,]

```

```
C10 <-C
C10[,,8,] <- apply(array(C[,,8:13,],dim=c(19,1,3,500)),c(1,2,4),sum)
C10 <- C10[,,1:8,]
C10 <- array(C10,dim=c(19,1,8,500))

#############################################################################
# Finally, we also multiply the catch by 2, and pretend that
# we have all the catch, not just the Norwegian.
#############################################################################

C10Total <- 2*C10
```

Appendix C

ChoosePriors

```
1
2 ##########################################################################
3 # Define mean natural mortality to be 0.18,                              #
4 # and calculate suitable hyper-parameters for the prior.  #
5 # Then plot and save the plot.                                           #
6 ##########################################################################
7
8 m=0.18
9 a <- 3
10 b <- a/m -a
11
12 pdf("C:\\Users\\Kjersti\\Google Drive\\Masteroppgaven\\Oppgaven_ny\\img\\
      Priorm.pdf", width=13, height=6)
13 x <- (0:10000)/10000
14 plot(x,dbeta(x,a,b), type="l", xlab="m", ylab="Density",col='blue')
15 xx <- c(x, rev(x))
16 yy <- c(rep(0,length(dbeta(x,a,b))),rev(dbeta(x,a,b)))
17 polygon(xx,yy,col="lightblue")
18 dev.off()
19
20 ##########################################################################
21 # Define mean fishery mortality for each age group,      #
22 # and calculate suitable hyper-parameters for the prior. #
23 # Then plot prior distributions and save the plot.       #
24 ##########################################################################
25
26 F <- c(0.005, 0.035, 0.1, 0.176,0.296,0.366,0.412,0.362)#ICES-estimates of F
27 f <- 1-e^(-F) # Transform to our fishery mortality.
```

```
a = c(1.01, 1.2, 1.5, 2, 2.5, 2.7, 2.9, 2.7)
b <- a/f-a

pdf("C:\\Users\\Kjersti\\Google Drive\\Masteroppgaven\\\\Oppgaven_ny\\img\\
    Priorf.pdf", height=7, width=10)
x <- (0:10000)/10000
par(mfrow=c(2,4))
for (i in 1:8){
plot(x,dbeta(x,a[i],b[i]), type="l", yaxt="n",xaxt="n",
      xlab=bquote("f"[.(i+2)]), ylab="",col='blue')
xx <- c(x, rev(x))
yy <- c(rep(0,length(dbeta(x,a[i],b[i]))),rev(dbeta(x,a[i],b[i])))
polygon(xx,yy,col="lightblue")
}
dev.off()

#############################################################################
# Prior for lambda. This is an example, when y=1985 and a=3. #
#############################################################################

ab <- 502000000   # Define mean
b <- 200000000    # Set a suitable beta
a <- ab/b      # Calculate alpha

pdf("C:\\Users\\Kjersti\\Google Drive\\Masteroppgaven\\Oppgaven_ny\\img\\1985.
    pdf", width=10, height=5)
x <- seq(0,2*ab,by=10000)
plot(x,dgamma(x,a,scale=b), type="l", xlab=expression(lambda["1985,3"]), ylab=
    "Density",col='blue')
xx <- c(x, rev(x))
yy <- c(rep(0,length(dgamma(x,a,scale=b))),rev(dgamma(x,a,scale=b)))
polygon(xx,yy,col="lightblue")
dev.off()
```

Appendix D

FindingQuantiles

```
1
2 FindQuantile <- function(Nb,w,q){
3   df <- as.data.frame(cbind("Nb"=Nb,"w"=w))
4   ordered <- df[order(df$Nb),]
5   weight.sum <-0
6   i <- 0
7
8   while (weight.sum < q){
9     i <- i+1
10    weight.sum <- weight.sum + ordered$w[i]
11  }
12  nq <- ordered$Nb[i]
13  return(nq)
14 }
```

Appendix E

SimulateDataSet

```
1
2 ######################################################################
3 # We start by defining the number of years (Y),      #
4 # seasons (S), ages (A) and number of samples (n).  #
5 ######################################################################
6
7 Y <- 19
8 S <- 1
9 A <- 8
10 n <- 50000
11
12 ######################################################################
13 # We assign the values we described in Chapter 6.      #
14 ######################################################################
15
16 q_sim <- array(1.5*10^-7,dim=c(Y,S,A))
17 f_sim <- c(0.005,0.035,0.1,0.176,0.296,0.366,0.412,0.362) # times 2.2 in Sec6
       .3.
18 m_sim <- rep(0.18,A)
19 lambda_sim <- rep(10^9,A)
20 for (a in 2:A){
21   lambda_sim[a] <- lambda_sim[a-1]*(1-m_sim[a-1])*(1-f_sim[a-1])
22 }
23
24 ######################################################
25 # Use the same values in every year.          #
26 ######################################################
27
```

```
f_sim <- aperm(array(f_sim,dim=c(A,Y,S)),c(2,3,1))
m_sim <- aperm(array(m_sim,dim=c(A,Y,S)),c(2,3,1))
lambda_sim <- aperm(array(lambda_sim,dim=c(A,Y,S)),c(2,3,1))

##############################################################################
# Obtain samples of abundances, catches and indices.                         #
# We use the normal distribution to approximate the                          #
# Poisson and binomial distributions if lambda is large.                     #
##############################################################################

if (max(lambda_sim)>= .Machine$interger.max){
N_sim <- array(rnorm(Y*S*A,lambda_sim,sqrt(lambda_sim)),dim=c(Y,S,A))
C_sim <- array(rnorm(Y*S*A*n,N_sim*f_sim*(1-m_sim),sqrt(N_sim*f_sim*(1-m_sim)*
    (1-f_sim*(1-m_sim)))),dim=c(Y,S,A,n))
Ind_sim <- array(rnorm(Y*A*100,N_sim*q_sim,sqrt(N_sim*q_sim)),dim=c(Y,A,100))
}
else{
N_sim <- array(rpois(Y*S*A,lambda_sim),dim=c(Y,S,A))
C_sim <- array(rbinom(Y*S*A*n,N_sim,f_sim*(1-m_sim)),dim=c(Y,S,A,n))
Ind_sim <- array(rbinom(Y*A*100,N_sim,q_sim),dim=c(Y,A,100))
}

##############################################################################
# Finally use round(), because the variables should be integers.             #
##############################################################################

N_sim <- round(N_sim)
C_sim <- round(C_sim)
Ind_sim <- round(Ind_sim)
```

Appendix F

StockSizeIS

```
1  StockSizeIS <- function(Ind, Cb, q){
2    ##############################################################
3    # Use C to define dimesions of Y, S, A and n. #
4    ##############################################################
5
6    Y <- dim(Cb)[1]
7    S <- dim(Cb)[2]
8    A <- dim(Cb)[3]
9    n <- dim(Cb)[4]
10
11   ###################################
12   # Estimate model parameters #
13   ###################################
14
15   ##################### Estimate lambda ######################
16   qlambda <- array(apply(Ind, c(1,2), mean),dim=c(Y,S,A))
17   lambda <- qlambda/q
18
19   # Force lambda to be larger than the catch.
20   for (a in 1:A){
21     for(y in 1:Y){
22       if(lambda[y,1,a]<max(Cb[y,1,a,])){
23         lambda[y,1,a]<- max(Cb[y,1,a,])+1
24         print("lambda was smaller than the catch!")
25       }
26     }
27   }
28
```

```
###### Estimate natural and fishery mortality (m and f) ######

meanCb <- apply(Cb,1:3,mean)
m = f <- array(0,dim=c(Y,S,A))

for (y in 1:(Y-1)){
  for (a in 1:(A-1)){
    Nbar <- lambda[y+1,1,a+1]+meanCb[y,1,a]
    f[y,1,a] <- meanCb[y,1,a]/Nbar
    #m[y,1,a] <- 1 - meanCb[y,1,a]/(lambda[y,1,a]*f[y,1,a])
    m[y,1,a] <- 1 - lambda[y+1,1,a+1]/(lambda[y,1,a]*(1-f[y,1,a]))
  }
}

m[Y,1,]<- m[Y-1,1,]
m[,1,A] <- m[,1,A-1]
f[Y,1,]<- f[Y-1,1,]
f[,1,A] <- f[,1,A-1]

#####################################################################
# Make array to store estimated numbers of fish,                    #
# and draw "starting values" for a=A.                               #
# Use normal approximation since lambda may be large.               #
#####################################################################

est.N <- array(0, dim=c(Y,S,A,n))
est.N[Y,S,,] <- round(array(rnorm(n*A,lambda[Y,S,],sqrt(lambda[Y,S,])), dim=
    c(A,n)))


#####################################################################
# Define the importance weights and draw samples of N|C,            #
# given the estimated model parameters.                             #
#####################################################################

weights <- array(1/n, dim=c(Y,n))

for (y in (Y-1):1){
  Nbar_ySAmin1 <- rbinom(n,round(est.N[y+1,1,A,]+Cb[y,S,A-1,]+Cb[y,S,A,]),
      lambda[y,S,A-1]*(1-m[y,S,A-1])/(lambda[y,S,A-1]*(1-m[y,S,A-1])+lambda[y
      ,S,A]*(1-m[y,S,A])))
```

```
67      Nbar_ySA <- est.N[y+1,1,A,] + Cb[y,S,A-1,] + Cb[y,S,A,] - Nbar_ySAmin1
68
69      est.N[y,S,A-1,] <- Nbar_ySAmin1 + rpois(n,lambda[y,S,A-1]*m[y,S,A-1])
70      est.N[y,S,A,] <-  Nbar_ySA + rpois(n,lambda[y,S,A]*m[y,S,A])
71
72      for (a in (A-1):2){
73        if(lambda[y,S,a-1]*m[y,S,a-1]<=.Machine$integer.max){
74          est.N[y,S,a-1,] <- est.N[y+1,1,a,] + Cb[y,S,a-1,] + rpois(n,lambda[y,S
              ,a-1]*m[y,S,a-1])
75        }
76        else{
77          est.N[y,S,a-1,] <- est.N[y+1,1,a,] + Cb[y,S,a-1,] + rnorm(n,lambda[y,S
              ,a-1]*m[y,S,a-1],sqrt(lambda[y,S,a-1]*m[y,S,a-1]))
78        }
79      }
80
81      #---------- Calculate importance weights ----------#
82      log.w <- apply(dbinom(round(apply(Ind[y,,],1,mean)),round(est.N[y,1,,]),q[
            y,1,],log=TRUE),2,sum)
83      weights[y,] <- exp(log.w-max(log.w))/sum(exp(log.w-max(log.w)))
84
85      print(paste("ESS for year", 1984+y))
86      print(1/sum(weights[y,]^2))
87    }
88
89    ##############################################
90    # Calculate total importance weights #
91    ##############################################
92
93    lw1 <- log(weights)
94    lw <- apply(lw1,2,sum)
95    w <- exp(lw-max(lw))/sum(exp(lw-max(lw)))
96
97    print("Total ESS")
98    print(1/sum(w^2))
99
100   ##########################################
101   # Define what the function returns. #
102   ##########################################
103
104   return(list("N"=est.N,"w"=w,"f"=f,"m"=m,"ess"=1/sum(w^2),"lambda"=lambda))
```

```
105 }
```

Appendix G

StockSizeISprior

```
1
2  StockSizeISprior <- function(Ind, Cb){
3    ############################################################
4    # Use C to define dimension of Y, S, A and n. #
5    ############################################################
6
7    Y <- dim(Cb)[1]
8    S <- dim(Cb)[2]
9    A <- dim(Cb)[3]
10   n <- dim(Cb)[4]
11
12   ############################################################
13   # Start drawing parameter values                           #
14   ############################################################
15
16   #-------------------m-------------------#
17   m <- array(rbeta(Y*S*A*n,3,13.67),dim=c(Y,S,A,n))
18   # m <- array(rbeta(Y*S*A*n,1,1),dim=c(Y,S,A,n))
19   # m <- array(0.18,dim=c(Y,S,A,n))
20
21   #-------------------f-------------------#
22   fangst <- 1-exp(-c(0.005, 0.035, 0.1, 0.176,0.296,0.366,0.412,0.362))
23   alpha = c(1.01, 1.2, 1.5, 2, 2.5, 2.7, 2.9, 2.7)
24   b<-(1-fangst)*alpha/fangst
25
26   f <- array(0,dim=c(Y,S,A,n))
27   for (a in 1:A){
28     f[,S,a,] <- array(rbeta(Y*n,alpha[a],b[a]),dim=c(Y,n))
```

```
}

f <- array(rbeta(Y*S*A*n,2,30),dim=c(Y,S,A,n))

#----------------------lambda-----------------#
lambda<- array(0, dim=c(Y,S,A,n))

ab = c(528732000, 323311000, 97995000, 47269000,20823000, 6481000, 3181000,
    969000)
beta = c(rep(5*10^7, 3),5*10^6,5*10^6, rep(5*10^5,3))
alpha = ab/beta

lambda[1,S,,] <- array(rgamma(A*n, alpha, scale=beta), dim=c(A,n))
lambda[,S,1,] <- array(rgamma(Y*n,502000000/(2*10^8) , scale=2*10^8),dim=c(Y
    ,n))

for (a in 2:A){
  for (y in 2:Y){
    lambda[y,S,a,] <- lambda[y-1,S,a-1,]*(1-m[y-1,S,A-1,])*(1-f[y-1,S,A-1,])
  }
}

#----------------------q----------------------#
q <- array(apply(Ind,c(1,2), mean)/apply(lambda,c(1,3), mean),dim=c(Y,S,A))

#########################################################################
# Make array to store estimated numbers of fish,       #
# and draw "starting values# for a=A.                  #
# Use normal approximation since lambda may be large. #
#########################################################################

est.N <- array(0, dim=c(Y,S,A,n))
est.N[Y,S,,] <- array(rnorm(n*A,lambda[Y,S,,],sqrt(lambda[Y,S,,])), dim=c(A,
    n))

#########################################################################
# Define the importance weights and draw samples of N|C,    #
# given the sampled model parameters.                       #
#########################################################################

weights <- rep(1/n,n)
```

```
67
68  for (y in (Y-1):1){
69    Nbar_ySAmin1 <- rbinom(n,round(est.N[y+1,1,A,]+Cb[y,S,A-1,]+Cb[y,S,A,]),
70                           lambda[y,S,A-1,]*(1-m[y,S,A-1,])/(lambda[y,S,A-1,]*
71                           (1-m[y,S,A-1,])+lambda[y,S,A,]*(1-m[y,S,A,])))
72    Nbar_ySA <- est.N[y+1,1,A,] + Cb[y,S,A-1,] + Cb[y,S,A,] - Nbar_ySAmin1
73
74    est.N[y,S,A-1,] <- Nbar_ySAmin1 + rpois(n,lambda[y,S,A-1,]*m[y,S,A-1,])
75    est.N[y,S,A,] <-  Nbar_ySA + rpois(n,lambda[y,S,A,]*m[y,S,A,])
76
77    for (a in (A-1):2){
78      if(max(lambda[y,S,a-1,]*m[y,S,a-1,])<=.Machine$integer.max){
79        est.N[y,S,a-1,] <- est.N[y+1,1,a,] + Cb[y,S,a-1,]
80        + rpois(n,lambda[y,S,a-1,]*m[y,S,a-1,])
81      }
82      else{
83        est.N[y,S,a-1,] <- est.N[y+1,1,a,] + Cb[y,S,a-1,]
84        + rnorm(n,lambda[y,S,a-1,]*m[y,S,a-1,],sqrt(lambda[y,S,a-1,]*m[y,S,a
              -1,]))
85      }
86    }
87
88    log.w <- log.w+apply(dbinom(round(apply(Ind[y,,],1,mean)),
89                                round(est.N[y,1,,]),q[y,1,],log=TRUE),2,sum)
90    weights <- exp(log.w-max(log.w))/sum(exp(log.w-max(log.w)))
91
92    print(paste("ESS for year", 1984+y))
93    print(1/sum(weights^2))
94
95    ################################################
96    # Resample if ESS is below some limit.  #
97    ################################################
98
99    if (1/sum(weights^2)<0.5*n){
100   resample <- as.vector(rmultinom(1,n,weights))
101   for (a in 1:A){
102      est.N[y,1,a,] <- rep(est.N[y,1,a,],resample)
103       Cb[y,1,a,] <- rep(Cb[y,1,a,],resample)
104      m[y,1,a,] <- rep(m[y,1,a,],resample)
105      f[y,1,a,] <- rep(f[y,1,a,],resample)
106      lambda[y,1,a,] <- rep(lambda[y,1,a,],resample)
```

```
107       }
108       weights <- rep(1/n,n)
109       log.w <- log(weights)
110       }
111     }
112
113     ##############################################
114     # Define what the function returns.  #
115     ##############################################
116
117     return(list("N"=est.N,"Cb"=Cb,"w"=weights,"f"=f,"m"=m,"ess"=1/sum(weights^2)
            ,"lambda"=lambda))
118 }
```